DOCUMENT RESUME

ED 120 253 TM 005 216

AUTHOR	Kiros, Fassil G.; And Others
TITLE	Educational Outcome Measurement in Developing Countries.
INSTITUTION	Georgetown Univ., Washington, D.C. Public Services Lab.; Haile Sellassie I Univ., Addis Ababa (Ethiopia).
SPONS AGENCY	Agency for International Development (Dept. of State), Washington, D.C.
PUB DATE	75
NOTE	198p.
EDRS PRICE	MF-$0.83 HC-$10.03 Plus Postage
DESCRIPTORS	Academic Achievement; Affective Tests; Cross Cultural Studies; *Developing Nations; Educational Assessment; Educational Development; Educational Economics; Educational Planning; Educational Policy; Educational Resources; *Input Output Analysis; Job Training; *Measurement; Nonformal Education; Program Evaluation; Resource Allocations; Social Action; Socioeconomic Status; Standardized Tests; Student Evaluation

ABSTRACT

Fiscal and other economic constraints facing developing nations underscore the need for careful and detailed planning for education focused on the identified purposes. Until recently educational requirements have been measured in terms of inputs. Effectiveness measurements change the basis for judging the educational system from such input or process variables to output indices that can yield more direct assessments of deficiencies, gaps, and additional requirements. This volume represents an initial effort to bring together information about educational outcomes, including types of measures, the role and use of measurement, and national policy consequences of measurements. The purpose is to provide a basis for Ministers of Finance, Planning, and Education to evaluate how and where they might best direct their educational systems to further educational goals and to stimulate national development. Sections discuss measuring academic achievement, attitudes and attributes, teacher grades and teacher tests, socioeconomic status and achievement, cross-national evaluation: the IEA Study, modernization: the social and political impact of education, nonformal education and functional competence, measuring work skills, economic outcomes of education, educational planning and evaluation, use of outcome measures in schools, improving resource allocation, and outcomes and evaluation. (RC)

Documents acquired by ERIC include many informal unpublished materials not available from other sources. ERIC makes every effort to obtain the best copy available. Nevertheless, items of marginal reproducibility are often encountered and this affects the quality of the microfiche and hardcopy reproductions ERIC makes available via the ERIC Document Reproduction Service (EDRS). EDRS is not responsible for the quality of the original document. Reproductions supplied by EDRS are the best that can be made from the original.

TM

ED120253

Educational Outcome Measurement in Developing Countries

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

TM005216

PUBLIC SERVICES LABORATORY
GEORGETOWN UNIVERSITY, U.S.A.
AND
THE INSTITUTE OF
DEVELOPMENT RESEARCH, ETHIOPIA

2

Educational Outcome Measurement in Developing Countries

Fassil G. Kiros
Institute of Development Research, Ethiopia
Selma J. Mushkin
Bradley B. Billings
Public Services Laboratory,
Georgetown University, U.S.A.

Public Services Laboratory
Georgetown University
Washington, D.C.
1975

The following individuals participated in the researching and writing for this volume.

Haile Sellassie I University

R. L. Jones*
Ayalew Kenno
Assefa Mehretu
Teshome Mulat
Lakew W. Tekle
Makonnen Yimer

Georgetown University

Ezra Glazer
Virginia Herman
Margaret Ladd
Raven Lidman
Stephen Stageberg

*Dr. Jones is presently at the University of California, Berkeley.

This book was produced as part of a project sponsored and funded under a contract with the Agency for International Development. Views expressed herein are those of the authors and are not necessarily those of the funding Agency or the universities with which the authors are associated.

Foreword

In March 1973 Public Services Laboratory of the School of Foreign Service of Georgetown University entered into a technical services contract with the Agency for International Development, State Department of the United States. The goal was to produce a volume on the measurement of educational outcomes in developing countries. What distinguished this effort from similar technical assistance projects in the past was the collaboration between Georgetown University and Haile Sellassie I University in Addis Ababa. As full partners in the endeavor, personnel at both universities worked together closely in producing this volume.

The measurement of the multiple educational products, out of both the formal school environment and the nonformal community, involves crossing many disciplinary lines. It involves drawing on the experiences of educators from a range of less developed countries. At Georgetown, the Public Services Laboratory carried main responsibility drawing on faculty of the School of Foreign Service, and the help of the Director of International Studies at Georgetown, Rev. Harold Bradley, S.J., and his staff. Haile Sellassie I University of Addis Ababa, Ethiopia, provided expert technical competence from several academic departments, including Psychology, Education, and Economics. The Institute of Development Research served as the focal point for the research and writing, and the University Testing Center, headed at the time by Dr. Reginald L. Jones, documented many of the testing procedures and problems encountered in a growing society. Beyond the explicit cooperation from University personnel, a further aid in the collaborative effort was the assistance provided by government officials through many of the government agencies dealing with education in Ethiopia.

During the first year of the joint project a preliminary research document was prepared. A subsequent international review conference was held in Addis Ababa in April 1974. Individuals with diverse backgrounds drawn from countries in Asia and Africa discussed the work and offered detailed suggestions for improvement. The present volume profited greatly from this review process. A companion handbook on the utility and uses of measurements is being considered for issuance under the auspices of the Organization of American States. This handbook is to be reviewed in Paipa, Colombia, in May 1975.

An important source of background material for this volume was a series of expert papers prepared by individuals drawn from many areas of education. These experts wrote on particular policy application problems of educational measurement, planning, and evaluation. They include: Charles S. Benson, J.C.B. Bigala, Charles L. Blaschke, Samuel Bowles, Russell G. Davis, Frank Farner, Herbert Gintis, Frederick H. Harbison, Torsten Husen, Nathan Kravetz, Donald M. Levine, Ernesto Schiefelbein, and Manuel Zymelman.

Throughout the course of this project, the greatest learning experience was derived from the interaction of individuals at Georgetown University and Haile Sellassie I University in Addis Ababa. Each group had a unique role to perform and a different insight into the effects of education in the developing world.

We want to thank all those who gave so willingly of their time to this cooperative international effort to clarify the means of measuring educational outcomes.

Peter F. Korgh
Dean of the School
of Foreign Service
Georgetown University

Contents

Foreword v

1 Introduction 1
2 Measuring Academic Achievement 10
3 The Other As: Attitudes and Attributes 26
4 Teacher Grades and Teacher Tests 45
5 Socioeconomic Status and Achievement 52
6 Cross-National Evaluation: The IEA Study 63
7 Modernization: The Social and Political Impact of Education 79
8 Nonformal Education and Functional Competence 95
9 Measuring Work Skills 105
10 The Economic Outcomes of Education 117
11 Educational Planning and Evaluation 139
12 Use of Outcome Measures in Schools 157
13 Improving Resource Allocation 167
14 Outcomes and Evaluation 178

Bibliography 185
Index 188

1

Introduction

Education in the developing world is playing a major, multi-purpose role —economic, social, and political—with diverse weights assigned to these purposes in different nations. In pursuit of these ends education in almost all nations has claimed a large share of resources. This share reflects the priority given by a nation to the development of its people's capacity to work and to advance toward a more elevated social and economic status.

Fiscal and other economic constraints facing developing nations are severe. Resources for education come at sizable opportunity costs in terms of optional resource use. These constraints underscore the need for careful and detailed planning for education focused on the identified purposes. A question central to planning is measurement of the progress made toward these purposes. How is effectiveness of education to be measured? How can we know what a nation gets for the resources spent on school systems? How can we know when education has well served a developmental project?

Until recently educational requirements have been measured in terms of inputs. Such indices as the proportion of persons in each age group enrolled in school, classroom size, expenditure per pupil, and teacher training and compensation were the operative criteria in judging an educational system. Effectiveness measurements change the basis for judging the educational system from such input or process variables to output indices that can yield more direct assessments of deficiencies, gaps, and additional requirements.

Major resource allocation and financial policies are dependent on output concepts and their measurement. When the objectives of educational systems are described and measured, the product of education can also be described and quantified. Such measurements can be used as guidelines for need assessment and for the allocation of resources to meet those needs. The measurements can also become basic data for determining the actual results of budget and program decisions, for evaluating such decisions and for designing methods of school finance. Investment in human capital, for example, is undertaken by nations to increase economic productivity and to enlarge the national product. It makes good sense for developing nations, especially those with many labor resources and relatively little physical capital, to develop and use those human resources effectively. But the effectiveness of such human development will depend upon whether the training produces skills that can be used productively in the economy.

Output measurements has as its framework input-output analysis. Such analysis and the theory of production from which it derives do *not* carry with them a system of values for determining goals in such human resources areas as education and health. "Output" is only an analytical construct, an empty term whose meaning must be filled in by its users. Those who wrestle with input-output analysis in education—or any other human resources area—must initially settle upon, through *some* rationale, a developmental output or set of outputs to be sought.

EFFECTIVENESS OF EDUCATION

Education policy is formulated daily by faculty, administrators, and parliamentary bodies. One might conclude, then, that there does exist some consistency in educational objectives. If a coherent set of educational objectives does exist, then it should be relatively easy to identify the outputs toward which all the activity and discussion are directed, and to develop criteria for measuring how well the goals are being met. But the *very first step* is to define educational goals and objectives. Then the goals and objectives inventoried might be classified according to similarity, priority, or some other criteria. That would be the kind of approach recommended by a program planner used to dealing in fields other than social resources. In education and other similar endeavors a neat input-output analysis is difficult.

Educational policy makers tend to be less than accurate in their statements of goals. Indeed the goals may sometimes be contradictory. Furthermore policy making is a continuous process subject to restatement, adjustment, redefinition, and reinterpretation. Consequently, inventories of policy objectives in a human resource field such as education may be unstable bases for the formulation of output measures. In any event getting such a systematic synthesis of objectives from all contributory sources often is not possible. The search for goals and objectives as a guide for selecting output measures in education may be a search for *criteria* on the basis of which alternative candidates for output measures can be identified.

In the absence of clear and consistent guidance from statements of goals and objectives, what criteria might be employed? Often input-output analysis of education as a developmental process turns to the calculation of the contribution of education to such economic variables as "economic growth" or GNP increases, or increases in productivity or skills created in defined occupations. While there can be little argument with economic growth (or its proxies) as *one* possible criterion (or set of criteria) of educational output, there also can be little argument that education contributes to a broad range of other developmental objectives. These should be assessed by a number of other criteria. Just as one can ask, "What is the contribution of education to economic growth?", one can also ask a range of equally significant parallel questions on other goals. For example, what is the contribution of education to social and national cohesion or stability? To social mobility? To cultural development? To civic and political participation? To an individual's sense of progress, modernity, and well-being? To an individual's sense of personal and social effectiveness? A substantial range of criteria should be invoked, for electing to measure the outputs of education by any single criterion involves a judgment. In short, in policy settings the so-called "criterion problem" is, at its heart, also a "value problem."

Furthermore, there does not exist a single metric in education comparable to market "price" by which all outputs can be commonly and summarily valued and compared. Because a single satisfactory index of educational product has not been developed, educational outputs must be defined in *multiple* dimensions. These dimensions, we suggest, relate to the multiple development targets of education—some involve personal development, some societal, and some are related to the quality of life. Before we turn to an analysis of those dimensions of educational output and how much more needs to be known before we can measure how well our educational institutions contribute to them, we should note briefly that educational institutions are not the only factors contributing to the three dimensions of development we mentioned above. Students enter school already the product of previous learning experiences. Therefore measurement of the outputs of education must be directed toward *changes* in characteristics—and only those changes occasioned by the educational system itself.

Similarly schools do not end a learning process. There is no finished product of the educational process. Individuals leave formal education at various times with differing skills, knowledges, and attitudes, but all can be changed by force of living, working, or traveling. Therefore to measure the changes in a school leaver's life as if the educational process was the sole factor in it would be distorting. To say that a graduate of a certain school or program achieves more, due to his education, may give higher marks of efficiency to the school's process (as opposed to the individual himself or the school's selection process) than is merited. However, a learner is the *output* of previous learning experiences, whether formal, nonformal, or informal.

Unlike most production processes, in education there is no *single* producer who controls the full range of inputs required for a given product. Instead the sources (inputs) of individual learning are multiple; the formal school represents only one. Other sources include, obviously, nonformal education and out of school experience in the family and with peers, in the larger community, and with the media, as well as out-of-classroom experiences during the formal school years. The individual utilizes his *own personal resources*, among them his time, abilities, acquired skills, physical and emotional energy to learn. It is in this sense that an individual is *both* a producer and a consumer of the outputs of education. This means that most measures of educational output—for example, of language development—will reflect the *combined* effects of multiple sources rather than the singular effects of any formal schooling.

These qualifying considerations indicate how difficult is the task of linking a particular outcome to formal education as the sole cause. Indeed in most instances formal education is not the sole and sufficient cause of any effect. It is only one of many interacting causes.

DIMENSIONS OF EDUCATIONAL OUTPUT

Investment in education serves broad purposes in a nation's social and political development; therefore measurement of outputs will reflect a wide range of educational objectives. Thus educational planning as a part of overall planning for economic development, social mobility, and political advancement has to be concerned with the questions of what programs, at what levels, for whom in the population, and with what timetable. Smooth economic growth requires a balanced growth of physical investment, the supply of educated manpower, and employment opportunities generated by new commercial and industrial investments. Without a careful synchronization of manpower development and industrialization, deficiences in education appear as bottlenecks to economic growth; but without a careful synchronization of education and employment, deficiencies in opportunities for employment appear as bottlenecks to social development.

Education as a means to *personal development* is complex and many-sided. Stated simply the problem is, "What personal attributes does a certain society and time feel should be developed through formal and non-formal education?" Certain aspects of personal development are emphasized today because of their special relevance to the problem of societal development and motivation for change. In medieval times in the West for instance, what education existed was geared to the preservation of biblical texts; during the renaissance, of Greek and Roman classics. Today's attributes—for some—include the capacity for invention, for innovation, and for development of the "modern" man with motivation for risk and entrepreneurship.

Schumpeter's work places great weight on individual initiative and attitudes fostering change as determinants of economic growth. A capacity to invent new forms and concepts plus skills of *innovation* are deemed worthy of cultivation because they are thought to foster the technological change that provides a spur to growth.

Instruments designed to measure such attributes are now available. The measures are variously captioned and differ in comprehensiveness and validity. But among the measures are these: capacity for risk taking, "modernity" indicators, creativity assessment, and characteristics for innovation. The fact that methods exist to test such changes in personality brought about by education indicates that other tests can be devised to measure practically any set of desired personality changes. For instance a society less interested in Western-type economic development and more interested in building or retaining a society of cultural and ecological preservation may opt to seek personality changes better suited to those goals than "entrepreneurship," and these changes can be measured as well. The educational structure itself is also a means toward developing a potential for "advance" or any other social goal.

The best developed measures of personal development due to education are in the areas of achievement and attitude testing. Over the years psychologists have developed batteries of tests which attempt to gauge the acquisition of skills and attitudes by the individual, and these tests have been adapted, with varying degrees of success, to the needs of the developing countries. While there are a variety of ways—some highly refined—to classify the variables which contribute to learning and thus to general behavior, there has appeared in the literature a *gross* distinction between "cognitive" factors on the one hand and "non-cognitive" or "affective" factors on the other. This distinction has found its way into the general writing on educational economics, evaluation and reform. According to frequent usage "cognitive factors" refer to "thinking" and to commonly recognized intellectual skills and processes. In contrast, the phrase "affective factors" refers to values, attitudes, interests, feelings, emotions, motivations, and personality attributes.

Because systematic and quantitative research on "cognitive" variables are customary parts of certification, attention to school tests, measurement, and curriculum development have been largely in this domain. The familiar tests of "cognitive" abilities (including tests of general "intelligence," aptitudes, and achievement, derived from early research) are now used, at least in developing nations, on a larger scale. Beyond the use of cognitive test results for measuring individual scholastic performance and for screening students—for example, from high school to college—these data represent the principal base on which education evaluation and analysis now rest.

Whatever their limitations as partial and imperfect measures of learning, achievement data such as national qualifying exams are used, in part, because they *exist*; supplementary or alternative relevant bodies of data do not. Systematic analyses of education would take on a substantially more varied and meaningful form if data on additional dimensions of individual learning and general development were available. Because "cognitive" testing is available and its data have value when properly used, a number of steps need to be taken to improve testing's utility and relevance. First, work is needed on the development of methods for establishing "equivalence" among the scores of the principal achievement tests to overcome some of the variability in the absolute value and meaning of separate scores derived from a range of tests of different design, scope, and effectiveness. Second, *new* cognitive measures (to reflect more fully an individual's level of language development or problem-solving ability, for example) need to be developed. Third, methods of standardization for population differences are needed to adjust existing achievement test data in order to reduce some of the "biases" in scores arising from differences in sex, income, and race. Despite the variety of steps being taken these efforts are directed primarily toward improving the relevance and utility of "cognitive" data.

Much more will be said about the measurement of cognitive skills, academic achievements, and the capacity of the educational system to ferret out talent and develop it to use for national development. But study suggests that there are other important aspects of education—the impression on personal attitudes and attributes. Education, formal, nonformal and informal, has an effect on an individual's developing self and helps determine whether a person feels controlled by fate or responsible for his or her own destiny. At the same time the individual's degree of self-esteem and sense of control, that is, confidence, will affect achievement in school.

In contrast to the steady concern with achievement variables, measures of other factors (such as affective and other non-cognitive variables) which contribute to school and life performance have been neglected. Consequently discussions of evaluation and analysis of education are sometimes premised on the odd notion of an abstract "cognitive" man who has no emotions, motivations, values, attitudes, interests, or personality characteristics to influence, guide, and direct his behavior. This naive view is being corrected formally by (a) social science research on a broad spectrum of variables beyond knowledge achievements; (b) empirical studies of the differences in the self-conception, habits, attitudes, and motivational dynamics of lower-class compared to upper- or middle-class children, which derive from different environments and childrearing practices; and (c) analyses of the role of "non-cognitive" factors which employ correlational and factor analytic techniques to elaborate the relationship between personality characteristics and school achievement. The latter studies are especially relevant to output measurement because they underscore the influence of variables *outside* the achievement domain on what have been traditionally considered by school people as "cognitive performances." In general they attempt to assess the relative contribution of academic achievement of tested ability or aptitude on the one hand, and of measured personality traits and motivations on the other.

The *quality of life* of a nation is in a sense a composite product of the knowledge, skills, and values of the people who compose it. Various components of knowledge are important to an understanding of the impact of such components on quality of life, and the impact of education on the productivity of the workers of a nation and the volume of goods and services that result. But two communities, or two nations, may be equal by such counts as productivity, income, national resources, or volume of goods and services available, yet the quality of life may differ sharply. In one community, for example, individuals may understand the processes of human reproduction and the consequences reproduction may have on crowding and limited resources. In another, the people may have no such understanding. Do the two communities have equal capacity to benefit from living? It is doubtful. Or, to cite another example, in one community there is understanding about nutritional values of such limited foodstuffs as are available and about how these limited resources can be used for better health and child development. In another, this knowledge may not exist. Life has a different quality in the two places. In addition, ill health attributable to lack of knowledge may result and this, in turn, may reduce production and exacerbate the already existing malnutrition. Even without the undesirable feedbacks of ignorance, the two communities have different levels of well being. Indeed "health" is a component of quality of living that ranks high in the scale of human values. Health involves more than nutrition, and many sides of health are dependent upon knowledge. An understanding about immunizations and disease, about the processes of contagion and infection, is essential to good health and disease prevention. And the general level of health depends upon the acquisition of knowledge about specific health practices. Still other aspects of living are critical to health maintenance. Knowledge about sanitation, waste disposal, water systems, storage and preservation of foods are all health related. And skills must be learned for converting abstract health knowledge into performance.

Knowledge of housing and construction making for homes as pleasant and safe as feasible is still another component influencing the quality of life. Values vary, but in each community there are standards and concepts of the acceptable and of beauty and comfort, as well as a range of knowledge and skill that can fulfill as closely as possible the aesthetic standards of that community.

A large number of specific measurements of required skills and knowledge can be enumerated to provide specific criteria for assessing educational effectiveness. We list a few here as illustrative of the level of knowledge:

- Percentage of water resources in a community used both for bathing and drinking,
- Percentage of families in which water is boiled prior to use,
- Percentage of births that are stillborn,
- Percentage of births that are attended,
- Percentage of women of child-bearing ages that have a choice about births,

- Percentage of children immunised, and
- Percentage of tuberculosis-infected persons attending very young children.

There are, of course, other factors relating to the quality of life besides health and sanitation that can be taught or instilled. Such qualities as appreciation of art and literature, or political awareness and activity, can also be measured to determine to what degree they are—or could be—outputs of education.

We can envision also the linkages between education and *societal development*. Formal education creates or develops learning capacity, creative talent, and the ability to adjust. These are important for specific job productivity as well as many other areas of societal development. Since it creates a fund of general usable knowledge which may include history, language skills and knowledge of the political process, and develops specific usable skills such as reading, writing and the ability to study, it contributes to flexibility and growth in work as well as living. Formal education also creates attitudes and attributes which bear on almost all facets of development including the characteristics of punctuality, orderliness, ability to take orders and organize work. Moreover, on the one hand it can help create cultural homogeneity and smooth over religious and ethnic differences, thereby reducing tension and providing the stability of communications and the distribution of knowledge needed for growth. Or on the other hand it can create the critical mass of educated people which many social scientists believe is required to break down traditional mores and create the environment necessary for diversity, experimentation, economic growth, and also political and social change. Nonformal and specific education providing specific skills—such as woodworking, plumbing, electrical, argicultural skills—and specific awareness—such as in health care and consumerism—can increase productivity, change traditional economic structures, and alter consumption goals and personal behavior.

As indicated earlier, the attitudes toward progress, toward change from tradition, and toward risk taking and modernization are vital to development.

THE PURPOSE OF THE STUDY

Increasingly, foreign assistance to developing countries is being tied to analytical studies that consider educational aims and content for human resource development. Through several organizations—International Bank of Reconstruction and Development (IBRD), United National Educational Scientific and Cultural Organization (UNESCO), International Institute for Educational Planning (IIEP), and Organization for Economic Cooperation and Development (OECD)—the donor nations have sought to encourage and support efforts to re-examine educational systems. They have urged the development of educational structures, curricula, teaching equipment, teaching methods, and teaching certification to meet the special problems and conditions in particular developing countries. Objective evaluation of ongoing educational methods has been emphasized, along with the analysis of new programs and activities.

When output data on educational programs and projects are available, gaps in knowledge, skill, attitudes in the population can be identified. Deficiencies, especially those deemed functional for the society, become the basis for issue definition and analysis. Such analysis can inform policy officials about the size and nature of the problem. In short, the quantifications derived from analyzing educational outputs set the stage for generation of policy and project options. The output measures could become a part of the documentation of need and, when combined with data on financial capacity, could serve to quantify external funds to be sought.

Indeed determination of outputs of education in a measurable way underlies improved management of resources. For instance unless the results of public outlays can be assessed, any determination of priorities among competing program claims may be the result of only subjective political decisions and may lead to inefficient use of public resources. The present volume represents an initial effort to bring together information about educational outcomes, including types of measures, the role and use of measurement, and national policy consequences of measurements. The purpose is to provide a basis for Ministers of Finance, Planning, and Education to evaluate how and where they might best direct their educational systems to further educational goals and to stimulate national development.

2

Measuring Academic Achievement

In the past planners have used general statistical indicators of educational needs in making decisions about resource allocation. Recently in developed countries attention has turned to the use of assessments of student attainment in defining system outputs and future needs. In this chapter we look at measures of student educational outcomes in the developing nations. We will define two general types of assessment and review the current state of educational assessment in developing countries—the characteristics of national examinations, the examination reform movement, and the growth of standardized testing in developing countries. The second part of the chapter focuses on decisions and problems confronting education ministers instituting standardized testing programs.

MAXIMUM AND TYPICAL PERFORMANCE

Educators have traditionally thought of learning outcomes as occurring in three broad categories:

Cognitive learning is the process by which we come to understand and organize the world as we perceive it. Cognitive skills are abilities such as acquiring knowledge, understanding phenomena, solving problems, creating new forms and ideas, and communicating effectively.

Affective learning involves the growth and maturation of aspects of the personality such as character and temperament. Many of the characteristics labeled "affective" in education are partly cognitive in nature—a purely affective characteristic might be a mood or temporary emotional state. More lasting characteristics such as values or attitudes are largely cognitive but have important affective consequences.

Psycho-motor, or perceptual-motor, learning is a crucial dimension of child development before age seven. This aspect of learning encompasses skills such as balance, left-right discrimination, hand-eye coordination, agility, endurance, flexibility, strength, relaxation, perception of body image.

However the more we learn about learning and development, the more these categories seem to overlap; abilities in each category are strongly interrelated. Certainly skills in each domain affect the development of faculties in the others. Much recent investigation has emphasized the mediating mechanisms in learning which have both cognitive and noncognitive components. These include motivation to achieve, expectancy of success, self-esteem, cognitive style, and inhibitors such as anxiety. These characteristics are both inputs and outcomes of education and have a profound effect on achievement.

A useful classification of outcomes from an assessment perspective distinguishes between capacity measures and those bearing on expected performance. One type of assessment is measurement of what a student can do—the maximum behavior or ability. This type of measurement may be concerned with what a student can learn (aptitude testing) or with the expression of that capacity in what a student has learned (achievement testing).

The second general type of assessment is the determination of characteristics which affect what a student will do—the typical behavior that can be expected. Appraisal of personality characteristics and their mediating mechanisms fall into this category, and will be discussed in Chapter 3. Current educational measurement research approaches summative and predictive assessment as a combination of aptitude/achievement testing and some optimal group of the second type of measures.

CURRENT EXAMINATION PRACTICES AND THEIR MODERNIZATION

Educational policymakers in developing nations are now examining existing assessment practice in the light of national goals as well as the current state of the art of testing. A corresponding inquiry into the connections between curriculum assessment and educational objectives—relationships long neglected—is gathering momentum. Worldwide, there has been a surge of interest in the reform of examination practices inherited from European traditions. The national external examination on the English or French model is now seen in many developing countries as largely irrelevant to present national needs and realities.

The external examination is often a written essay or oral test based largely on the recitation of memorized material. It evolved in post-renaissance Europe as a mechanism of control over the schools by subsidizing agencies (church or state) and was exported to the developing world with the European educational structure. Today the external examination is still designed and controlled by agencies outside the school and often beyond the influence of teachers or administrators. Typically the examinations have tended to function quite independently of educational policies guiding the schools. Yet because they determine admission to higher education, the examinations exert an enormous sway over teaching. However studies have discounted the validity of European external examinations as selection devices, noting variations in scoring and level of difficulty. These external school-leaving examinations are under special scrutiny in developing nations where they often control the content of instruction, dominate the attention of students and teachers, and cause widespread anxiety and frustration.

The need for basic changes is clear. The traditional examination does not produce comparable and replicable data about what students know and do not know. It does not assess achievement, but measures linguistic fluency and certain elusive intellectual qualities associated with a humanistic and classical education. The continued emphasis on these European classical traditions in the developing countries may lead to an impractical and irrelevant educational experience.

Even when national examinations are revised, as they have been in the last decade in many countries, usually they are not subjected to the rigorous analysis required to ensure validity. A case in point is the Ethiopian School Leaving Certificate Examination (ESLCE) which has been given to students in the last year of academic secondary school for the last 19 years. The ESLCE is prepared and scored by subject matter specialists associated with the public schools, the Ministry of Education and Haile Sellassie I University. A candidate who takes the ESLCE is successful if he attains five passes at C level, including the Amharic language examination, the English language examination, and the mathematics examination. There is no limit to the number of times a candidate can take the ESLCE, and the time allocated for each examination varies.

The ESLCE has long been controversial. The percentage of those passing the exam sank from 89 percent in 1950 to 20 percent in 1966, causing widespread public dissatisfaction and disputes between teachers and examination officials over the cause of the high failure rate. In 1967, a number of commissions were set up to examine the content of the examinations and the high school curriculum. Participants declared that contributing to the failure rate were promotion of students without adequate skills from the lower grades, poor instruction by high school teachers, and inadequate student preparation for the examination. Teachers claimed they had ambiguous questions, limited coverage of subject matter, and allowed too little time for answering questions.

At the conclusion of a five-day session, the commissions recommended enrichment of the curriculum and revisions of the ESLCE, including the substitution of objective test items for some essay questions. But no significant reduction has been made in the number of failures, which varied between 83 percent and 74 percent between 1969 and 1973.

The issue to be addressed is whether the revised ESLCE is valid (does it accurately test the knowledge, skills, and understanding that a high school graduate should have) and reliable (are the results consistent).

Preliminary findings have indicated that the predictive validity of the ESLCE for university performance has been modest and the reliability of the test has never been established. Yet the examination is used as the yardstick for determining "successful" completion of secondary school.

STANDARDIZED TESTING: ONE DIRECTION OF REFORM

The widespread examination reform movement, stimulated by extensive unhappiness with existing examination practices, will have a great effect on future collection of data on educational outcomes. Proposed solutions range from the abolition of all exams to the substitution of standardized tests. Some educators are seeking to develop new essay-type examinations with shorter questions to provide a better sampling of subject matter. The major trend, however, is toward objective standardized testing, which is repeatable for various populations at different times because the test content (though not the exact items), procedure, and scoring are fixed.

A few examples illustrate the trend toward standardized tests in developing nations:

Singapore's Ministry of Education, working with the Cambridge Overseas Examination Syndicate, instituted standardized testing programs at the end of the sixth year of primary school, the fourth year of secondary school, and the second year of postsecondary school. The Primary School Leaving Examination covers two languages (choices include English, Chinese, Malay or Tamil), mathematics and science.

In Vietnam, the Ministry of Education, Culture and Youth abolished the eleventh grade baccalaureate in 1973. The ministry is now designing standardized tests, to be administered to high school students completing the twelfth grade in 1974. These tests will replace the twelfth grade baccalaureate.

The Philippines, through a private Philippine organization, administered a standardized college entrance examination (CET) on a voluntary basis two years ago. The testing organization has now contracted with the government to develop a national college entrance examination (NCEE) for admission to all postsecondary schools. The NCEE will have two parts: a subject proficiency section measuring English, mathematics and science; and a mental ability section consisting of tests in verbal analysis, number and letter series, word-number relations, and abstract reasoning.

Indonesia provides an interesting example of testing reform resulting in both localization of control and national data gathering. In the 1950s education officials concluded that the current examinations and curriculum were unsuitable for a democratic educational system: the uniform curricula did not reflect the cultural and geographic diversity of Indonesia, and the harsh external examinations were not functional for selection. The government abolished external examinations, following a nationwide upgrading of teacher skills in testing and evaluation, and decentralized the examination organization. After 1969 all examinations were produced either by teachers or, in some parts of the country, by provincial committees. Beginning in 1973 each school developed its own examinations.

At the same time the ministry developed a standardized testing program for educational planning purposes. These tests will be given to all children in grades one through 12, and cover mathematics, science, social science, and Indonesian languages. The questions are short answer and include the cognitive areas of knowledge, comprehension, and application. Two functions of testing are thus clearly separated: the evaluative and guidance function for students and teachers, and the assessment of education outcomes for educational planners.

MEASURING EDUCATIONAL OUTCOMES WITH STANDARDIZED TESTS

The necessity for assessment of educational needs and outcomes is common to all countries. But developing countries have a smaller margin for experimentation in the allocation of resources. It is in these countries, where the need for a close analysis of investments in education is greatest, that the data base for decisionmaking is least complete and accurate.

Planners must be able to weigh the relative merits of program investments with regard to quantitative increase in education, needed changes in teacher recruitment and training, needed changes in curriculum, and the diversification of the institutional role of education in national life. Viable planning in these areas requires measurements of the numbers of people to reach each level of education annually, the availability of the resources for educational development, and the abilities of individuals inside and outside the formal educational system.

It is the third type of measurement that is most deficient and it is this type of data which standardized testing can supply. Specific data on the acquired skills and educational deficits of demographically differentiated groups of students can be collected on a continuing basis for program analysis and planning. Furthermore basic skills of large numbers of people outside the formal system—adults and youth who never entered school or who dropped out—can be assessed for vocational placement or by allowing them to reenter formal schooling at the proper level. Finally standardized testing is appropriate for certain types of nonformal programs. If technical skills are taught in diverse ways, certificates of program completion may have little meaning in the market place. Although practical tests will continue to be important, a well designed and validated standardized test can assess large numbers of applicants economically (see Chapter 9 on work skill assessment).

Implementation of standardized testing is a sizable task. The Education Minister or official must first decide whether to use an existing test—published in another country or by the International Association for the Evaluation of Education (IEA)—or design a new test for that country alone or cooperatively with other countries. Test construction is a considerable undertaking; it is costly and requires specialized expertise and time for development. For this reason the use of existing tests is a serious option to be considered. However the investment in test development may well be justified for long-term planning and educational benefits. Countries need testing instruments designed in the context of their particular social needs and problems. Both options will be examined.

USE OF EXISTING TESTS

Tests developed in other countries are rarely appropriate for use in developing countries without adaptation. Even where the language of the test is a desired feature some reading level adjustment may be needed. An exception might be the International Assessment (IEA, discussed in Chapter 6) instruments, which were designed for cross-national use and have been administered in some developing nations. The advantage of the IEA tests is that they were painstakingly constructed and extensively validated. But it must be remembered that the International Assessment effort was initiated to analyze across nations the effects of various in-school and out-of-school variables on learning. The research orientation of these tests on international comparisons may not coincide with the outcome information needs of a country.

The advantages of adaptation of existing tests are economy and the relative speed of the process, compared to test construction. Furthermore a standardized test has little value if it has not gone through the processes of standardization and validation discussed later in this chapter. These procedures may present great problems to developing countries in terms of knowledgeable personnel, data processing systems and financial support, and hence use of existing, previously validated tests may offer a solution to these problems.

However, the problems with adaptation may outweigh the initial convenience and economy. There are several basic considerations in test adaptation: language, content, mechanics, and rationale.

Translation is a self-explanatory task but it may pose difficulties if the test contains vernacular phrases or proverbs.

Content adjustments are routinely made by test adapters on a superficial level (for example, changing monetary or measurement systems), but other issues of content are often ignored. The accuracy of a test is greater if the applicants are familiar with the examples and illustrations used. In a vocational aptitude test a common error has been to assume that an item measuring a certain type of mechanical aptitude must be presented in terms of the western machine or tool, rather than a more familiar tool from the local environment. But the test is intended to measure the target skill and knowledge of the machine to be used on the job.

Testing mechanics must be modified by the test administrator according to circumstance. Test-taking is a skill improved by practice and familiarity. A separate answer sheet sometimes causes confusion for groups unaccustomed to standardized tests. Another element is timing: cross-cultural research shows that people have very different senses of timing. The time limit prescribed by a standardized test may have to be extended or eliminated to be fair to some groups of subjects. The instruction to "work as quickly and as accurately as you can" involves a complex type of judgment which is particularly alien to many cultures.

The test rationale deserves careful consideration. Every test is designed within a particular cultural context and is usually validated against practical criteria in that particular culture. The consequences of translating the test to another culture can be evaluated only in terms of the purpose of the test. Poor language facility, slow work habits and lack of abstract thinking will affect test scores, but may also affect the criterion behavior which the test was designed to measure (for example, in a vocational aptitude test). In other cases this is not necessarily true.

To summarize, there are at least six levels of adjustment necessary for adaptation of instructional materials and examinations:

- translation,
- vocabulary changes for the proper reading level,
- illustration and photograph substitutions, to reflect local conditions,
- adjustment in procedures—implied or specified—to match the expectations and experience of the learners,
- content adjustment to reflect local culture and life style, and
- accommodations to the learning styles of the students.

The last three are complex and require intimate knowledge of the culture. These considerations underline the necessity for nations to acquire training psychometricians with the technical and cultural requisites for analyzing and designing test modifications for the functional requirements of the educational system.

The Haile Sellassie I University Testing Center in Ethiopia has had extensive experience in using and adapting tests such as the Differential Aptitude Test, the Davis Reading Test, and the Amharic Otis 100 (adapted from the Otis Quick Scoring Mental Ability Test).

Some of the conclusions drawn from the experiences of the Testing Center in assessing educational outcomes with adapted measures are:

1) The studies conducted demonstrated that cultural factors greatly affect the ability of test items to elicit the intended responses. Comprehensible test items must be constructed in the light of the examinee's cultural background.
2) Since lack of familiarity with the procedures and types of tests poses a threat to the validity of the instrument and reliability of the results, it is essential that adequate information about tests be given to examinees ahead of time.
3) The indiscriminate use of standardized tests from developed nations is likely to be loaded with sampling error due to differences in the populations.
4) Validation (the process of ensuring accuracy) for the criterion behavior toward which the test is directed must be undertaken in the

adapting country. Before tests are used as predictors of success for the desired criterion, a scientifically designed pilot must precede to verify that the tests actually predict the intended performance of the invididual. This aspect of test development needs greater emphasis for better validity and reliability of the test results.

DEVELOPMENT OF A NEW TEST

Designing new standardized tests rather than adapting tests from other nations has several distinct advantages:

1) The effect of cultural influences on test performance may be pronounced. An adapted test may not be suited to the population because of built-in assumptions about cognitive skills, familiarity with concepts or phenomena, or cognitive style. Necessary revisions in test mechanics and administration may be extensive.
2) A new test can be designed specifically for the curriculum and educational objectives of the formal school system. Tailor-made tests are essential for assessment of many objectives in nonformal programs.
3) The test can be designed to supply the particular kinds of outcome data for the particular populations (grade levels) required by planners.
4) The process of test development is an opportunity to increase levels of awareness and expertise with educational measurement among teachers, administrators, and educational researchers.

The high cost of test development can be minimized by a regional pooling of financial and manpower resources among several geographically and culturally related countries. Measurement experts, teachers, educational psychologists, anthropologists and educational planners working together in a centralized research laboratory can generate common examinations and specialized tests for the participating countries. The West African Examinations Council is an example of such an arrangement.

If a central problem is the lack of measurement specialists in a country or region, a solution is to bring testing experts from abroad to coordinate the test development project. The contract with the consulting organization can specify training in measurement and test design for personnel in the Ministry of Education and workshops for local administrators and teachers in measurement, test item writing and evaluation. It has proven important, in the developed countries, to have teachers involved in some phase of test design to promote a supportive attitude toward the procedure and rationale of standardized testing.

IMPLEMENTING A TESTING PROGRAM

In this section we examine, step-by-step, the basic decision areas facing an Education Minister or regional policy group engaged in implementing a testing program. These include: 1) defining the purpose of the test; 2) determining grade level and timing of test administration; 3) emphasizing aptitude, achievement testing or a combination of the two; 4) choosing norm-referenced or criterion-referenced instruments; 5) defining the basic educational objectives and content to be covered; and 6) test construction and validation.

Defining the Purpose of the Test

The reason for giving a standardized test must be clearly understood by test administrators, school personnel, and the students. In countries where such testing is common, dissatisfaction sometimes results from imprecise definition of the goals of the test which, in turn, determine the appropriate methodology, the analysis of the data, and the fairness of the practical consequences. A test which is designed for one purpose is usually inappropriate and often unfair when used for a different purpose.

Educational measurement provides either needs assessment or treatment adequacy assessment data. Needs assessment is identifying the specific goals toward which education should be directed. Sometimes needs assessment is used to identify specific learner needs in terms of understood goals, for prescriptive or placement purposes. Treatment adequacy assessment refers to evaluation of the educational process; it is either formative (on-going testing during the instructional program) or summative (takes place at the end of the program to judge its success). These three types of testing yield antecedent, formative, and outcome data. In some cases tests at an intermediate level may be used both as a summative evaluation of a completed program and for needs assessment for treatment planning for the next level. But in principle the purposes are distinct and ideally the tests to serve each should be tailored to the purpose.

Some specific uses for outcome measurement are noted below.

- students and teachers—diagnosis of the learner's strengths and weaknesses for individual guidance and evaluation of student progress;
- administrators—evaluation of educational programs, selection of students for the next level of education, distribution of students in different types of programs;
- employers—certification of competence in required skills and knowledge;
- educational planners—evaluation of the quality, efficiency and adequacy of the educational system as a whole; and
- educational research—contributions to the development of educational learning theory.

Although one test can serve more than one function (for example data on individual student achievement for students and teachers as well as gross data on achievement for national planners), it must be carefully designed to do so. Tests cannot accomplish many goals indiscriminately; they must be designed for specific purposes in the context of the particular situation and goals of a country. This constitutes one of the difficulties of transplanting tests designed for other countries.

Grade Level and Timing of Tests

An early decision that must be made is which grades are to be tested and with what frequency. A logical schedule is testing at the end of each stage: lower and upper primary, intermediate, and secondary. Another most costly plan is to test every other grade, beginning in the second grade. This may be appropriate where elementary grades are grouped in units of two grades each, as in the Arab states. On the other hand testing too often—for example, at the end of each school year for admission to the next grade—can have damaging effects on students. Again the certification and system evaluation functions of testing must be kept clearly defferentiated.

Yearly testing for admission to the next grade is expensive, may cause extensive anxiety, and has a low predictive value for young children who develop cognitive capacity at different rates. The lock-step approach to instruction and individual assessment handicaps children who develop more slowly—a child should not be labeled a failure after only one or two years of school. In the early years of schooling it is important to balance standardized tests with more subjective teacher evaluations. At this stage standardized tests are more useful for group assessment than for individual assessment.

Aptitude and Achievement Testing

Educators developing a new system of assessment to replace the traditional examination have a choice between aptitude testing; a combination of aptitude and achievement testing; and achievement testing alone. This choice is affected by the potential uses of the test. Our primary concern in this report is the group assessment of all incremental gains by students—what might be called the value-added concept in measurement. This clearly requires achievement testing on a regular basis so that longitudinal data can be collected.

However, planners in some countries may wish to implement standardized testing not merely for planning purposes but also for purposes of selection for high levels of education. Combining two functions of assessment—for planning purposes and for selection—cannot be taken lightly since the former implies achievement testing and the latter has in the past implied aptitude testing. A brief review of the advantages and disadvantagers of the options listed above in the light of both these functions follow.

Aptitude testing attempts to measure innate ability; it has been favored for predicting future performance. A few key areas (such as verbal and mathematical ability) can be emphasized, and therefore the aptitude test may be more economical than an achievement test. The aptitude test also theoretically circumvents problems and unequal quality and differing curricula in schools and consequent unequal academic achievement of students.

The disadvantages of aptitude testing are two-fold. First, it does not yield systematic data for national evaluation and planning. Second, it is extremely difficult to design a scholastic aptitude test that does not to some degree measure achievement or environmental characteristics. Indeed general aptitude or I.Q. tests purporting to measure intelligence, or basic ability to learn, have been the subject of much controversy for decades. Factors such as socioeconomic status (and hence greater opportunity to learn) are so highly correlated with intelligence test scores that the tests have acquired some disrepute, especially in a cross-cultural context. The obvious deficiencies of some intelligence tests spurred an effort to create "culture fair" tests which substitute abstract concepts for universal pictorial symbols for culturally-linked content. However further research revealed that even geometric shapes are perceived differently in different cultures. It is now felt that after eliminating all culturally-differentiating material the remaining content may be too trivial to be valid for any practical criterion.

Specific scholastic aptitude tests are based on the theory that mastery of a specific skill results not from generalized ability but from particular abilities which can be isolated. These mental abilities, or traits, were identified by factor analysis of responses to test items by large numbers of people. Scholastic aptitude tests are discrete tests generated from the clusters of related items resulting from the factor analysis (There are also nonscholastic aptitude tests which are widely used: skill tests for abilities such as hand-eye coordination, and vocational aptitude tests simulating particular job skills).

Some researchers objected to this model, in which the mind consists of independent constituents which can be separately measured, proposing an alternate model of the mind as an organic structure of complex interrelationships. They maintain that specific cognitive abilities, such as performing mathematical operations, consist mainly of the command of the knowledge pertaining to the process and that general ability is primarily clusters of closely related specific abilities. Therefore much of what has been seen as innate ability would actually be attributable to previous learning experiences.

The validity and usefulness of the mental trait theory to explain individual differences in learning is yet to be resolved. Aptitude tests have a diagnostic function—they can be used to prescribe appropriate learning environments for individual students (for example, programmed instruction, lecture method or unstructured environment). But it seems clear that most aptitude tests to some extent measure achievement or environment.

Combining aptitude and achievement testing has been proposed by Krug and Schwarz (**42**, 1972),* who argue for secondary school selection in developing nations by means of a two-tiered process. First, an achievement test would screen out students who had not mastered skills necessary for success in the secondary course; then a scholastic aptitude test would be used as the predictive instrument to identify the best candidates for secondary school. The rationale is based on the assumption that the past schooling experience of test takers is unequal in developing countries, and hence achievement alone is an inequitable criterion.

---

* The boldface number refers the reader to the bibliography for publishing details.

This proposal does guarantee that students selected for secondary school would possess the requisite skills for progress, but the use of an aptitude test as the ultimate predictor has the major drawback that it is difficult to design a "pure" scholastic aptitude test not actually measuring achievement. There is no way to know to what degree aptitude testing used in limiting educational opportunities may unfairly handicap part of a population.

It appears that the only way to assure fairness and gain detailed outcome data on educational outcomes would be to implement both aptitude and achievement testing programs simultaneously, which is not an economical procedure.

Achievement testing is a straightforward measure of educational outcomes. For students and teachers such tests provide feedback information on progress and problem areas; for administrators, they also allow comparative evaluations of programs or schools with a region. For the national planner, whose problem is to conduct educational assessment that will evaluate the effectiveness of the entire system, achievement testing is the logical tool. Aptitude tests, which are more selective in content and are based on the premise of prediction rather than of assessment, clearly will not satisfy the needs of planners or administrators for system evaluation. Achievement tests appear to have the capacity to perform both functions. Valid and relevant assessment requires tests based on educational objectives, and with adequate coverage. Coverage must concern itself not merely with sufficient sampling of curricula but with expected and desired behavior in terms of national goals; otherwise objective assessment is open to the criticism of trivializing the aims of education.

The difficulties imposed by differing curricula and quality of school instruction may be partially overcome by some recent technical developments in test construction, particularly the criterion-referenced test discussed below.

Selection of Instruments

Most standardized tests in use today are norm-referenced; they are designed to show how an individual scores in relation to others taking the same test. This test has been used widely in competitive situations.

In recent years testing experts have developed the criterion-referenced test (or domain-referenced test), designed to measure a student's ability to perform a particular skill or mastery of a concept. The significance of the criterion-referenced test is that it is informative about the breadth of measured knowledge but less informative in comparing individuals. Statistical considerations, as well as content validity, influence the inclusion of test items in the norm-referenced test, so that variability in scores will result. Criterion-referenced tests are constructed around specific learner objectives and could reveal that all of the students mastered the objectives or at the other extreme, that no one mastered the curriculum content sought. But if desired the criterion-referenced test can be designed to produce the kind of variability, or ranking of individuals, given by a norm-referenced test.

For nearly every function of testing criterion-referenced tests provide better information. Where the emphasis is on assessment of educational quality, they are clearly superior—although typically more expensive to produce than norm-referenced tests. Work is now proceeding on the technical problems of criterion-referencing. As models for these tests become more readily available they will become more economical and more applicable for selection. A test designed to assess mastery of skills needed to succeed in secondary school would be a better predictor of academic success in secondary school than a norm-referenced test based on the elementary school curriculum. The latter is more likely to yield misleading estimates of performance and to reflect existing social status in selection. Following a criterion-referenced screening test, candidates could be selected from a pool of qualified candidates according to other criteria (geographic, for example), resulting in greater access to education for nonprivileged students. Other factors which can be utilized as predictors of academic success, such as achievement motivation, are discussed in Chapter 7.

In most cases a carefully designed and administered standardized achievement test, possibly combined with other methods of evaluation, will probably be the best way to select students for further schooling.

Defining Objectives and Test Content

One reason for caution in adapting existing instruments has already been discussed: a good test, like an effective curriculum, reflects the values and educational objectives of the society in which it is given. Ultimately the aims of a society are the basis for evaluation of the educational system.

Objectives-based assessment entails close analysis of the cognitive skills implied by educational goals. There are three very general types of goals. Students must acquire the necessary knowledge and skills to a) live happily and successfully; b) fill social roles, including the nation's manpower goals; and c) preserve and develop the traditions and values of the society and humanity at large. The elaboration of these goals into specific skills and knowledge is the foundation of curriculum and evaluation design.

More specifically, Benjamin Bloom (5, 1956) has classified educational objectives in the cognitive domain in a hierarchical taxonomy with six ascending levels:

1) knowledge of facts, concepts, generalizations and methods;
2) comprehension, interpretation, and extrapolation of information;
3) application of principles;
4) analysis of material, recognizing relations and organization;
5) synthesis of parts to produce a new communication; and
6) evaluation or judgment.

A taxonomy of this type can help ensure that tests incorporate the higher orders of cognitive learning and not merely the lowest level of repetition of facts. Interpretation and reporting of test scores determine the degree to which program evaluation is possible. Tests should report a separate score for each objective rather than an average score.

Evaluation has a feedback effect on the instructional process. If examinations determine acceptance to higher levels of education, they tend to control the content of curriculum. It is counterproductive to promote national priorities in curriculum while neglecting the relationship between the content of examinations and national goals.

Test Construction

The construction of a standardized test involves many procedures including:

1) Writing to test specifications—a complete and explicit listing of all characteristics of the test. This means resolving all questions of test content, length, difficulty, scoring. All the important decisions about the purpose of the test and the educational objectives which it covers must be made at this initial stage.
2) Writing the test items—subject matter experts and test technicians often work collaboratively to write items within the framework of the specified values and objectives.
3) Pretesting the items—gathering and analyzing data on a sample response to the items.
4) Designing preliminary test forms.
5) Pretesting the preliminary test forms—for reliability, difficulty, time limits and so forth.
6) Designing the final test forms.
7) Administering the test to standardize and validate it.
8) Preparing a test manual and other materials.
9) Printing, publication and copyright.

Test accuracy is of vital importance in test construction. A basic question about tests is what they really measure. Good test instruments have three qualities: validity, reliability, and usability. The most important quality is validity, the extent to which the test measures what it is supposed to measure. It can be determined by statistical comparison of the measurements to some outside criterion, or by other means. If no convenient and measurable criterion is available, expert opinions may be used, which obviously introduces an element of subjectivity.

Content validity, important for achievement tests, is the extent to which the test includes a representative sample of the universe of content and objectives for the area being measured. The author of a national achievement test must attain content validity for varying curricula, complicating the task of achieving a close fit between instructional material and the test. An item by item analysis is necessary, either in designing or adapting an achievement test.

Construct validity is important for tests in the social/emotional domain. A construct is an abstraction—a conception underlying a test reflecting an intangible attribute—for example, anxiety. Anxiety as a construct cannot be observed directly. Its presence is inferred on the basis of certain observable events—for example, increased heart rate, palms sweating, and so forth. Consequently a test of anxiety may be said to possess construct validity to the extent that high or low scores on it predict some observable behavior or event which on logical or theoretical grounds might be expected to be a consequence of the test-measured state.

Concurrent validity is established by relating the test to a more established measure of the same attribute administered at approximately the same setting.

It is important to remember that no test is universally valid; it is valid for a particular time and circumstance.

Reliability refers to the consistency of measurement. All things equal, the score of a person on a test today should be the same as his score at some subsequent date, if no circumstances have intervened which markedly affect the score (for example, intensive study of questions missed or severe physical or emotional trauma). High reliability is essential if there is to be confidence in test scores.

Reliability is generally assessed in one of several ways: through the test-retest method in which a test is administered on successive occasions following an interval of time; using comparable split-halves of the test administered at the same time; or on the basis of an analysis of the interval consistency of test items. Correlational and other techniques for computing reliability coefficients are widely available.

Usability pertains to administrability, scorability, economy, format and the adequacy of norms. Determining usability is in the domain of the policymaker as well as the testing expert.

Some of the processes involved in test construction in a developing country are illustrated by the Ethiopian Oral Amharic Proficiency Test. The Haile Sellassie I University Test Center in Ethiopia has experimented with construction of new instruments as well as adaptation of existing ones. The Oral Amharic Proficiency Test, a test of proficiency in listening and understanding of spoken Amharic, was developed in the 1967-68 academic year. The items consisted of common sayings, word usages, word meanings, commonsense interpretation of statements, and so forth. Each item has three possible answers: True or False or neither True nor False. Some sample items are: "the best way to get from one side of the road to the other is to take a taxi"; "a window and a door do not mean the same thing"; and "if two individuals are equal in height then they are also equal in weight."

Out of a pool of 500 items, only 127 items were used in the final test. The rest were dropped because of item difficulty, ambiguity, cultural bias, and so forth, which became evident after repeated administration of the test to different groups.

The correlation between the oral Amharic scores and first and second semester grades ranged from 0.36 to 0.49. If taken as evidence of oral ability (for which the test was designed), the above figures are encouragingly high; its purpose was to measure proficiency in comprehension at an elementary level.

The following are some of the problems that were encountered in the development of the oral Amharic test: 1) insincerity of some trial examinees; 2) reluctance of essential faculty reviewers to make "critical" comments; 3) avoidance of culturally biased items; 4) construction of simple clearcut statements which are either true or false or neither true nor false; and 5) identification of difficult and/or ambiguous items.

CONCLUSION

The use of standardized testing programs to provide systematic data on the outcomes of schooling is a relatively new concept, but one that will be increasingly important in system-wide planning. Objective standardized tests have great potential value for facilitating educational decisionmaking on a national level. Measuring the effects of instruction is far from an exact science, but it can yield direct information on the quality of efficiency of formal and nonformal educational systems.

The task ahead for planners is to encourage the development of indigenous expertise in test construction and administration by establishing national or regional programs of test development. These programs, already ongoing in some areas, may well prove to be cost-effective. Traditional examinations often are not adequate for either planning or selection purposes in the expanding educational systems of developing countries; and the limitations of adapted tests may soon make them unacceptable to nationally oriented educators.

3

The Other As: Attitudes and Attributes

In Chapter 2 we discussed the first two of the four As of educational outcomes, aptitudes and achievements. In this chapter we consider attitudes and attributes, or characteristics with affective consequences for students. We shall consider the significance of attitudes and attributes in developing nations and some approaches to assessment of two important outcomes: self-esteem and locus of control, using the experience of the University Testing Center, Haile Sellassie I University, in measuring these outcomes.

Assessment of personal and social outcomes presents a new set of problems for the educator. Goals in the areas of attitudes, attributes, values and feelings are difficult to define, and it is even more difficult to achieve consensus on which should be adopted and which should be emphasized. To assess progress toward any goal in this area is technically complex and beset by the same ambiguities and political controversy of definition and assigning priorities.

The ambiguity inherent in analyzing, teaching and evaluating nonacademic skills does not diminish their importance. People in most countries, when asked what the goals of education should be, will stress outcomes such as loyalty and nationalism, civic participation, moral and character development, appreciation of culture and learning. But these goals, even when explicitly stated, are not easily translated into quantifiable objectives for teachers and students.

Leaders in developing nations recognize the attitudinal consequences of education as evidenced by goals framed by education ministers in many regional conferences in the last few decades. For example a major recommendation of the Commission on Primary Education for the Conference on Education and Scientific and Technical Training in Nairobi (1968) was that African governments should ensure that the primary education system of African countries contribute to the strengthening of national unity; bring about the social and cultural integration of children in the community; and act as factors of change and of economic and social development.

Policy statements from this and other conferences show sensitivity to the detrimental attitudinal effects of uprooting children from their families and cultures. In *Learning To Be,* Faure notes that the schools "inculcate values into school-children which estrange them from their surroundings, feeding intellectual and material ambitions which are becoming harder and harder to realize in a rural setting. Schools thereby push young people out toward the towns, having failed to instill in them the kind of values which should make people attached to their everyday surroundings" (Faure, **30**, 1972).

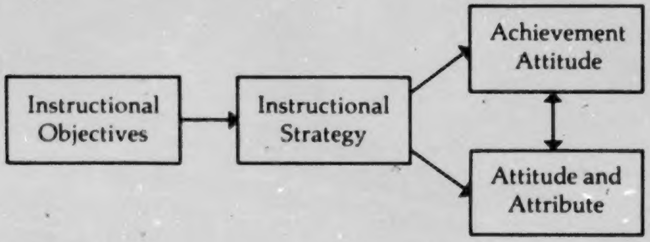
A balance must be struck between retaining vital cultural wellsprings and cultivating new patterns. These concerns are not contradictory, if nations can learn to absorb changes without creating new but equally rigid institutions. An Arab educator states: "We therefore have to see that our education produces citizens who are well equipped in science and technology, who are resourceful, adaptable, adaptive, responsible, realistic, efficient, enthusiastic, effective, and conscious of their problems and the way out of them. The concepts, the attitudes, the skills and the modes of behaving acquired in the past have to undergo changes for the present and more profound changes for the future" (El-Koussy, **20**, 1973).

IMPLICATIONS OF SCHOOLING FOR PERSONAL AND SOCIAL GROWTH

Education in traditional societies can be highly functional. Some societies still have formal learning structures in which observation of teachers plays a major role. The students learn military training, music, geography, religion, and language, as well as skills connected with their future roles—farming, hunting, housekeeping. A common characteristic of this kind of education is that the process of learning is emotionally charged and personalized. The knowledge of the society is highly valued and is transmitted individually or in small groups by teachers with high social status.

Gradually, as schooling became more institutionalized, the course of study became progressively less congruent with the skills and knowledge learned informally in daily life, more abstract and verbal, and less concerned with the moral and emotional lives of children. In many parts of the world the schools represented an oppressive foreign culture and actively denigrated the culture of the children, often with disastrous results for the individual's self-perception. (Such results are also reported by studies of the self-concepts of minority group children in the United States.)

It should be clear then that outcome assessment for planning and reform must treat many aspects of the lives of children in schools. Curriculum, teaching techniques, administrative patterns and other factors in formal and nonformal education structures have important effects on student development in nonacademic areas. Some of these effects will be described in later chapters in connection with the feedback effects of outcome measurement for curriculum, teaching and administration. It is important to note the reciprocal influences between achievements that are academic and other outcomes. Attitudes and values affect learning and are altered by it. The diagram below represents an instructional paradigm in which instructional objectives are phrased (in measurable terms), an instructional strategy devised for achievement of the objectives, and the objectives achieved are measured.



ATTITUDES AND ATTRIBUTES AS INPUTS TO EDUCATION

Attitudes and attributes function both as inputs and outputs of education. A child's attitude toward mathematics may greatly influence learning and performance in that area. At the same time success or failure in mathematics will influence the child's self-concept. In the past more attention has been directed at personal-social attributes as inputs. Much research in educational psychology has been oriented toward discovering relationships or correlations between personality variables and achievement, with the idea that psychological tests might be used as predictors of academic success. Some of the personality factors that have been found to have a positive relationship to academic achievement include independence, impulse control, achievement motivation, ability to make consistent judgments, persistence, order, endurance and stability. But the research results have not been entirely consistent with respect to the achievement implications of most personality variables. Many studies have found no relationship or a weak relationship between achievement and general adjustment as measured by the major psychological batteries. Yet we cannot conclude that personality factors are not crucial to learning, because the wealth of research studies indicate that they are. We simply cannot point to particular variables as the causal factors.

There are several possible explanations for the lack of clearcut findings. One is that personality factors interact in ways we do not wholly understand, and therefore studying characteristics in isolation is likely to be misleading. Even where multivariate or multiple factors studies were employed relationships tended to be weak (and one must remember that a relationship or correlation between two variables does not mean that a cause and effect relation exists). The studies often implicitly assume that the individual operates in a social vacuum; it may be that the social setting of the performance is an important variable.

Another problem is that many of the instruments used in this research were developed by clinical psychologists for the diagnostic identification of maladjusted individuals. These tests tend to be slanted toward negative personality traits or neurotic disorders; they often are quite inappropriate for educational assessment. And since what constitutes "maladjustment" is culturally determined, developing countries are particularly cautious about using foreign psychological tests.

In the last few decades educational psychologists have developed many instruments to measure particular attributes in a school setting. However their research or orientation makes them impractical for broader assessments because of basic orientation, dated or inappropriate items, impractical mechanics, and so forth. Published tests and others discussed in the literature were often designed by investigators with limited financial support, and therefore have not been subjected to the extensive and rigorous procedures which produce somewhat more reliable achievement and aptitude tests. In developing nations there is the additional problem of translation and cultural adaptation of published instruments; these are even more complex than in the case of a cognitive test.

ASSESSMENT OF ATTITUDES AND ATTRIBUTES AS EDUCATIONAL OUTCOMES

Many educators have devoted much attention to the importance of the affective domain. Benjamin Bloom and two associates designed in 1964 a classification model for human behavioral outcomes in the affective domain (Krathwohl, **41**, 1964). Choosing internalization as the ordering principle for all possible human responses, they describe a hierarchy of five broad categories of response to stimuli:

level 1.0 Receiving (the learner is conscious of a phenomenon or object and is inclined to pay attention to it).
2.0 Responding (the learner actively and voluntarily attends to or complies with a phenomenon).
3.0 Valuing (the learner believes that a thing, phenomenon or behavior has worth).
4.0 Organization (the learner begins to build an organized value system).
5.0 Characterization by a Value or Value Complex (the individual acts and responds in terms of his/her system of attitudes and values, or philosophy of life).

The authors have described each category and the subcategories in each by including sample educational objectives illustrating development on that level.

Other educators have concentrated on developing programs or intervention strategies to further personal and social growth by students. Some techniques include group and individual counseling, play therapy and dramatic activities, reinforcement approaches in a group setting, curriculum changes (for example, to focus on material students can identify with) or methodological changes (for example, structuring the instruction so that there is a high probability of success for the student).

But despite some promising programs assessment of affective student outcomes of education has been largely neglected. Noncognitive objectives are rarely included in instructional programs. Consequently their assessment as desired outcomes is at best an unsystematic enterprise, often limited to evaluations of scattered special programs. Ordinarily in the process of teaching such objectives tend to fall by the wayside and organized assessment is forgotten. Even when teachers are consciously aware of their affective objectives for their students, they often feel that they can best evaluate these outcomes subjectively—by observing their students' behavior. In many nations, another factor is the reluctance of educators to evaluate personality factors in students. In the last few decades it has been assumed by many that noncognitive development was an automatic by-product of cognitive growth, though this has not been convincingly demonstrated. There is great interest at present in specifying nonacademic goals and tracing nonacademic development, but this interest is stymied by the lack of standardized instruments.

An educator wishing to implement outcome assessment of attitudes and attributes faces a twofold problem. First, how is one to choose the significant variables or a particular variable from the vast array of psychological and social attributes that have been identified and measured? Second, how is one to locate good instruments to measure these variables? Chart 1 represents a framework or analytical structure of potentially relevant personal attitudes and attributes. This general classification scheme permits differentiation between broad types of characteristics.

Personal-social characteristics are particular personality traits which are specific individual expressions of the polarities which form the general dimensions of personality. Controlling mechanisms are the individual's style or manner of dealing with or organizing phenomena, and represent a fusion of cognitive abilities and personality factors. These characteristics, often called "cognitive styles" or "cognitive controls," include, for instance, field dependence, which is a person's ability to differentiate an object from its background. A test for field dependence is the rod and frame test, which measures a subject's ability to hold a rod perpendicular to a frame which is not perpendicular to the ground. Performance on this task is seen to predict the individual's approach to a broad range of tasks—it indicates whether phenomena are experienced as discrete from the field or background in which they are contained, or whether the subject responds globally to the dominant organization of the situation.

The third category includes values (beliefs or commitments; general attitudes that are fully internalized), attitudes (responses made to socially

Chart 1. Cognitive and Personal-Social Style

I. *Basic Drives*, e.g., self assertion, fear, gregariousness, succorance, curiosity

II *General Dimensions of Personality*, e.g.,

withdrawn vs. involved
masculine vs. feminine
rebellious vs. compliant
expressive vs. restrained

tense vs. relaxed
sensitive to others vs.
  self-centered
submissive vs. dominant
active vs. passive

apathetic vs. energetic
solitary vs. social
assertive vs. timid
aimless vs. purposeful

rigid vs. flexible
happy vs. unhappy
academically motivated
vs. otherwise motivated

III. *Areas of Personality Expression*, e.g.,

A. *Personal-Social Characteristics*

achievement
affiliation
aggression
anxiety
autonomy/independence
curiosity
deference
dependency
dominance
nurturance
creativity

B. *Controlling Mechanisms*

scanning-focusing
field articulation
conceptual differentiation
tolerance for delay of reward
intenal vs. external locus
  of control
risk-taking strategy
impulsivity-reflectivity
conceptual style
cognitive complexity
distractability
coping styles

C. *Values, Attitudes and Interests*

Interests: attitudes toward
manipulative cognitive aesthetic

1. family
2. groups (political, ethnic)
3. community
4. self
5. tasks (physical)
6. important beliefs and acts
7. national and international
8. nature
9. the future/posterity

Adapted from *Disadvantaged Children and their First School Experiences*: ETS-Head Start Longitudinal Study. Volume 1, Princeton, New Jersey: Educational Testing Service, August 1970.

significant elements in the individual's environment) and interests (the manifestation of an attitude in an individual's activities). While the general dimensions of personality are universally shared, the expressions of those dimensions are in part culturally determined; the most concrete category of interests may be quite localized (interests reflecting local institutions, geography and climate, and cultural activities).

Educational researchers generally investigate the categories of personal-social characteristics and controlling mechanisms. On the classroom level, teachers are more concerned with the third category of values, attitudes, and interests, which are the focus of the Bloom taxonomy discussed earlier.

On the level of system-wide assessment of educational outcomes, the significant noncognitive factors in all three categories listed here should be identified, using criteria derived from the educational goals of the nation.

A group of characteristics felt to be important in national development (for example, autonomy/independence, locus of control, self-esteem, attitudes toward community and nation) could be selected for the development of assessment measures to be applied longitudinally. These standardized test instruments, applied in standardized testing programs, will help determine whether the school system is reaching its goals.

Some aspects of this process are detailed below.

SELECTION OF MEASUREMENT INSTRUMENTS

We focus here on "pad and pencil" tests, the most efficient and economical method for wide-scale measurement of the qualities of personal development judged important to a nation. Other effective techniques such as observation, individual and group interview, or projective techniques, may be useful on a classroom or program level.

Most existing paper and pencil tests use the self-description mode of assessment. Description by students of their past and current experience in a structured format is the most fruitful method for systematic group assessment required for analysis and evaluation of psychological characteristics, cognitive controls, and attitudes, values, and interests. This category includes devices such as questionnaires, checklists, self-ratings, biographical inventories and attitude scales. As noted before, very few existing tests are directly applicable to school situations in developing nations —they lack validating support, are sometimes difficult to acquire, and are ambiguous in interpretation. However, they are the prototypes for the tests that should be developed for use in schools. Nations must apply resources to designing acceptable instruments for measuring those social outcomes which they judge to be important.

Several guides have been published in the United States which describe and evaluate published tests in the personal-social development area.

These include:

Buros' *Mental Measurements Yearbook*: The most comprehensive source of test information, the Yearbook includes availability and administration information, a list of published references to the instrument and often a detailed critical review. The seventh edition, published in 1972, contained listings for 1,157 tests and reviews of 798.

Robinson—ISR Series: Robinson and colleagues at the Institute for Social Research, University of Michigan, have published three volumes of test descriptions, *Measures of Political Attitudes*, *Measures of Occupational Attitudes*, and *Measures of Social Psychological Attitudes*. The last was revised in 1973 and includes descriptions and evaluations, with sample items, of 126 scales and measures.

CSE Test Evaluations: The Center for the Study of Evaluation of the University of California has published several books reviewing tests for school children according to detailed criteria. Two of the volumes are: *Elementary School Test Evaluations* (1970) and *Preschool/Kindergarten Test Evaluations* (1971).

Other sources for information are test publishers' catalogues, the quarterly *Test Collection Bulletin* of the Educational Testing Service in Princeton, New Jersey, measurement textbooks, abstracts and indexes such as *Psychological Abstracts*, journals and periodicals relating to educational measurement, and doctoral dissertations.

MEASURES USED IN OUTCOME STUDIES IN ETHIOPIA

As examples of measures of attitudes and attributes now available, we will look at a few scales administered by the Haile Sellassie I University Testing Center. Some considerations in adapting measures of attitudes and attributes will be considered later.

Attitudinal measures must be selected in the light of the particular culture in which the measures are to be used and the particular outcomes to be derived. A few measures were seen as potentially useful in Ethiopia, using the criteria of: availability, ease of modifiability, potential for illuminating possible relationships between input variables and educational outcomes, and usefulness for prediction of university performance. These characteristics are also related to qualities implied by modernization (see Chapter 7). The Testing Center modified existing tests for use with graduating high school seniors. Examples of tests and measures administered by the Testing Center were the following:

1. *Anomie*. A state of cultural and personal disorganization results when the individual is unable to refer his behavior and that of his associates to any stable set of standards. The individual caught in such circumstances is said to be alienated. Although not typically included in educational outcome assessments, measures of anomie could be included since

one goal of education is to reduce alienation and increase societal participation.

Sample items: It's hardly fair to bring children into the world the way things look for the future.

These days a person doesn't really know on whom he can count.

(All items required an agree/disagree response.)

2. *Status concern*. Attitudes toward status and mobility are measured by the value the individual places on symbols of status and on the attainment of higher status. Low responses on this measure would indicate achievement of a desired educational outcome of nonconformity.

Sample items: The raising of one's social position is one of the most imtant goals in life..

Before joining any civic or political organization it is usually important to find out whether it has the backing of people who have achieved a respected social position.

3. *Test anxiety*. Anxiety created by having to take tests is measured because there is evidence of a negative correlation between high text anxiety and test performance, although a moderate level of anxiety may facilitate performance. Given the importance attached to tests and test performance in Ethiopia, test anxiety is an important construct. Studies related to educational outcome assessment, therefore, must consider the impact of test anxiety on test performance.

Sample items: If examinations could be done away with, I think I would actually learn more.

Thoughts of doing poorly interfere with my performance on tests.

4. *Social desirability*. This methodological scale, used to study the validity of the various test and scale items, measures the respondents' tendency to describe themselves in socially desirable terms. If correlations between test item and social desirability scores are not low or negligible, it is impossible to separate responses to the test item from the respondents' tendency to place themselves in a favorable light. Items for which this is the case must be eliminated from any test or scale. The measure of social desirability modified from use at HSIU was that developed by Crowne and Marlowe (16, 1964) and consists of 28 items.

5. *Response set*—the tendency to respond to a test item in a manner which is independent of item content—contributes to the invalidity of test items. The Crowne-Marlowe Scale provided a measure for control of social desirability response sets. A second type of set contributing to the

invalidity of item response is the tendency to agree to a test or scale item independent of content (acquiescent or agreement response set). This tendency is controlled by wording scale items so that an approximately equal number are phrased in the agree and disagree directions.

Sample items: I have never intensely disliked anyone.

No matter who I'm talking to, I'm always a good listener.

SELF-CONCEPT AND LOCUS OF CONTROL

Two constructs in particular—self-concept and locus of control—are seen by many investigators as central variables. These two characteristics encompass the individual's view of himself and his view of the world in relation to him. We will examine the constructs and some assessment instruments for each.

Self-Concept

Self-concept (one's view of oneself) is a central feature in the personal world of every individual. Although a complex subject for study and assessment, it has the advantage of reflecting a number of other variables such as motivation, needs, values, and attitudes. Human beings tend to attach meanings and values to things, people, and events; the values attached to the self evolve in the course of interaction with the external world and in turn guide the behavior of the individual. Self-concept is thus a dynamic personality trait, changing with growth and experience. A practically identical trait often measured is self-esteem, which means liking and respect for oneself, or personal judgment of worthiness. Another highly related construct (personality trait) is self-acceptance.

The importance of self-esteem comes from its pervasive effect on behavior. The way people feel about themselves colors their perceptions of the world and their responses to it. Esteem attributes are more subject to perceptual distortion than physical attributes like height or wearing glasses. And since self-concept is private and invisible, it may lead to misinterpretations of motivations and intentions by others. For example, if a boy feels that he is a bad reader he will see reading aloud as a threatening and painful experience and will avoid it. A teacher might interpret this as a "bad attitude" or laziness. The teacher's reaction may further lower the student's concept of himself as a student and reader.

Dimensions of Self-Concept

Throughout this century behavioral scientists have produced numerous research studies and developmental theories concerned with self-concept. Some of the internal dimensions, or subselves, that have been described are the identity self (one's basic perception of self-identity, which influences and is influenced by one's behavior); the judging self (the self functioning as observer, standard setter and evaluator); and the ideal self (one's image of how one would like to be).

There are external dimensions as well—self-concept varies according to frame of reference. One might have different images of the self as student, citizen, worker, or family member. Self-concept as a whole is a composite of the esteem attached to all the subselves. A student may have a low concept of himself as a student but high self-esteem as a family member. Everyone's behavior varies from one situation to another, but a well-integrated person will have a consistent feeling of worth and self-regard while realizing that his abilities in different areas vary. Self theorists see self-actualized people as those who are inner-directed, motivated by personal goals and not the need for approval of others, and consequently lead rich and effective lives.

Results of Research

Self-esteem research has been complicated by the fact that it is difficult to correlate the data from the hundreds of studies done. Researchers have tended to design their own tests rather than use existing ones, and have focused on specialized problems so that broad conclusions are difficult to formulate.

Many studies have examined the discrepancy between the self-image and the "ideal" self-image of a person. It was originally hypothesized that a high disparity indicated maladjustment, but United States studies have shown that the disparity tends to increase with age, IQ, and cognitive complexity, among normal well-adjusted populations. Evidently older and more intelligent people in the United States have higher expectations of themselves and better defined perceptions of goals. Cross-cultural studies have indicated the influence of the cultural environment on this discrepancy. One study comparing American and Indian 12-year-olds, both with high socioeconomic status, found a much higher congruence between real and ideal self-concepts among the Indian children, particularly the Indian girls (Swart and Swart, **63**, 1970).

It has been found that a generalized high self-concept has a positive but weak relation to intelligence, and a stronger positive relationship to specific self-concept of school ability and reported school grades (Bachman, **4**, 1970). The importance of the opinions of "significant others" to self-concept has often been noted. Investigators have found a high correlation between the child's self-concept and the teachers' reports of perceptions of them, and also between the children's self-perceptions and their parents' perceptions of them. Other studies indicate that self-esteem is related to sex role identity for adolescents.

Researchers have also found a positive relation between self-esteem and academic success, but it is not a simple one-to-one correspondence since many other variables intrude. Coopersmith found that a high discrepancy between goals and performance, or a high discrepancy between self and teacher evaluations, led to repeated failure for 5th and 6th graders (15, p. 130).

Another general finding is that self-esteem is affected by the experiences of the individual. In United States studies, unsuccessful candidates for office and university students carrying heavy work loads lost self-esteem (71, p. 442). On the other hand, one study found that the self-esteem of black students increased after outstanding black speakers were brought to the school during a school year.

The self-concept is formed in the context of a particular cultural milieu and a particular peer group and it may be relatively independent of social and economic status. Some studies in the United States demonstrated that disadvantaged black children (defined as those coming from low income families living in low rent or subsidized housing) had higher self perceptions than advantaged children (62, 1971).

Measurement of Self-Esteem

Problems abound in the measurement of an attribute such as self-esteem which is idiosyncratic and not directly observable. Most self-theorists feel that it can be best appraised through self-reports. Self-report formats that can be used effectively for group assessment are forced choice scales (test takers are presented with two or three statements about something and must choose one) and Likert-type scales (the test takers specify whether they strongly agree, agree, are undecided, disagree or strongly disagree with a statement). Another format is the adjective check list in which 300 or so adjectives are listed and the respondent checks the ones felt to be self-descriptive.

Some sample self-esteem test items are (items answered either "like me" or "unlike me"):

"I often wish I were someone else."
"If I have something to say, I usually say it."
"I often get discouraged at what I am doing."

A problem with testing self-concept that has received some attention is the difficulty of including all the possible important dimensions of self-concept in a standardized test. A test may contain items about school and other situations without referring to areas from which individuals may derive high self-esteem—for example, unusual hobbies or activities. Some self-concept scales are generalized, while others are specific to the school situation. Educators must decide if they are interested primarily in the students' academic self-esteem or if they need more generalized information.

Locus of Control

In recent years an increasingly productive area of psychological research and measurement has been the investigation of personal cognitive styles of controlling mechanisms (affective mechanisms, in educator terminology). This style can be defined as the individual preferred modes of perceiving stimuli from the environment and organizing a response.

A particular control characteristic, which has been identified as important to individual development and on a broader scale to national development, is locus of control, sometimes referred to as IE, for "internal-external" control.

The concept behind this construct is that people view events which affect them as resulting from their own behavior or as controlled by outside forces: fate, chance, or powers beyond one's control or understanding. People who tend to see a cause and effect relationship between their own actions and things that happen to them are called "internal"; they feel that some control lies in themselves. People who believe that most of the things that happen to them are caused by external forces are called "external."

Many behavioral scientists believe this characteristic greatly affects a person's life, as well as the character of a society. A feeling of control over one's environment probably facilitates change, mobility and growth within that environment. A feeling of lack of control may lead to acceptance of the status quo, hopelessness, and frustration. The amount of effort invested in any activity is a function not only of motivation, but also expectancy of success; both of these elements are tapped in locus of control measurement.

Like self-concept, locus of control is a complex attribute. A person who attributes all responsibility to himself would be deeply frustrated in a situation that really is beyond his control: for example, being confined as a prisoner of war. In such a situation it would be more sensible to have an external orientation. Similarly it is risky to make comparisons of locus of control attributes across cultures or between different socioeconomic or ethnic groups. The realities of power vary widely between societies, and a healthy individual's outlook on life is tempered by reality.

However if not interpreted too rigidly, locus of control is a meaningful dimension of behavior and response to education and adaptability in later life. It is probably in the interest of most nations to promote an internal orientation among its citizens for social and economic development.

Research

Most research indicates that people are hindered by external locus of control beliefs. Achievement is depressed, health and safety precautions are less likely to be taken, delayed rewards are less valued, conformity and docility are higher. High achievement tests scores were found to be positively related to internal locus of control. In a large scale project, composite achievement reading, math and language grade placement, and academic self-concept were positively related to internal control for boys (Martin, **48**, 1972). A positive correlation between internal control and achievement has also been found for girls who scored low on social desirability scales, in other words, who were not dependent on others for approval (Nowicki and Walker, **52**, 1974). In studies of adults "internal" workers show improved job performance and more highly developed personal qualities relating to employability and job success than "external" workers.

The various measures described above (and others) have been used in studies carried out by the Haile Sellassie I University Testing Center. The results are briefly summarized in Chapter 7 and hence will not be reviewed here.

Variations or Types

Some dimensions of locus of control identified by factor analyses of pools of items from locus of control scales have been labeled: control ideology, or the subject's belief about the extent to which people have control generally; personal control, or the extent to which the subject believes in personal control; and system modifiability, referring to control over world affairs, war, racial discrimination, and so forth.

Another aspect of the construct is the application of a positive or negative value by the individual to his orientation—one might view the world as controlling and malevolent or controlling and benevolent. Another question that has been raised concerns linking fate, chance and "powerful others" together as external forces. One researcher has devised a scale which separated powerful people and chance in scoring for external control (Levenson, **43**, 1972). Another distinction that has been made relates to differentiation between control over positive and negative events. Some locus of control tests are designed to yield both scores. There is a great difference between a person who accepts responsibility for success and denies responsibility for failure and a person who follows the opposite pattern.

Despite the complexity of the construct, researchers in the United States have had considerable success measuring IE. Some are attempting to develop unidimensional scales. Some are specific to control over, and responsibility for, academic success and failure.

The formats of paper and pencil tests include forced choices between statements, the yes-no response to questions, and Likert-type format. Several non-verbal formats with cartoon pictures have been developed for children. These must be read aloud by the test administrator. Some examples of a locus of control item are:

When you do well on a test at school, is it more likely to be:

A. Because you studied for it, or
B. Because the test was especially easy?

Do you believe that most children are just born good at sports?

Do you think it is better to be smart than to be lucky?

The significance of the self-concept and locus of control constructs must be analyzed in terms of the present cultural milieu of a society as well as its future needs and goals. There are many unanswered questions about the meaning of personality variables in different societies. For example what is the relationship between self-concept and locus of control within different status groups? One hypothesis is that in a highly traditional society members of lower classes may have a high level of self-esteem (derived from the security of a class structure with clearly defined roles and expectation for individuals) combined with an external locus of control orientation that breeds fatalism. This configuration might be reversed for upper class members, who have greater mobility but higher self-expectations. Another problem to be more fully investigated is the effect of modernization, and the requirements of modernization, in relation to these variables. Generally it can be said that high self-esteem and internal locus of control are desirable outcomes; but much remains to be learned about the dynamics of these characteristics for different groups, that is, rural people, farmers, urban workers, women, shopkeepers, within the developing context.

THE ETHIOPIAN EXPERIENCE WITH TESTING ATTITUDES AND ATTRIBUTES

The Haile Sellassie I University Testing Center has experimented for several years with construction and modification of attitudes and attributes for use in Ethiopia.

Adapting instruments developed in other countries is less than ideal, but it is not always possible or necessary to develop new instruments for each situation. Yet it should not be assumed that tests and scales can be taken from other cultures without a critical appraisal of their cultural and psychometric properties.

The Testing Center has modified tests and measures developed in the United States by eliminating or rephrasing items seen to be culturally inappropriate. Future studies may take the form of adding items of similar content but designed to reflect the local situation.

Test development and modification requires a thorough understanding of the constructs underlying the test or measure to be modified. Those undertaking such work should be well informed about the culture in which the scale is to be used. An Ethiopian educational psychologist and an Ethiopian cultural anthropologist screened and modified the scales used in studies conducted by the Testing Center. Each test item was reviewed by the judges for its clarity and relevance to Ethiopian culture.

Item Appraisal in Operation

The items below were judged to be inappropriate within the Ethiopian context and then eliminated or modified. The process is necessarily subjective and will vary from one situation to another.

Items Eliminated:

The average citizen can have an influence on government decisions.

This world is run by a few people in power, and there is not much the little guy can do about it. (Internal-External Control)

With enough effort we can wipe out political corruption.

It is difficult for people to have much control over the things politicians do in office. (Internal-External Control)

Before voting I thoroughly investigate the qualifications of all the candidates. (Social Desirability)

My table manners at home are as good as when I eat in a restaurant. (Social Desirability)

If I could get in a movie without paying for it and be sure I was not seen, I would probably do it. (Social Desirability)

Within the context in which these test questions were administered (for example, as part of the Aptitude Test battery for university admission given in all provinces of the empire), political questions were deemed inappropriate, lest it be perceived that student responses to such questions would influence university admission. A byproduct of the removal of certain political questions from the Internal-External Scale was to improve its unidimensionality. Certain data indicate that the scale has several components, among them a political one. When the political items are removed the factorial structure of the instrument is improved.

The social desirability items described above were eliminated because of their unsuitability for utilization throughout the empire. For example, the question concerning getting into a movie without paying would be appropriate for use in Addis Ababa where there are many movie houses, but inappropriate in many other communities where there are none. The same rationale was applied to the question concerning eating in a restaurant. In addition to the obvious class bias associated with eating in a restaurant, restaurants are more likely to be available in the larger population centers than in the small communities, thus rendering it an unsuitable test item.

Items Modified:

I always try to practice what I preach. (Original Social Desirability item)

I always try to do what I tell others to do. (Modified item)

Although people sometimes compliment me, I feel that I do not really deserve the compliment. (Original Self-concept item)

Although people sometimes praise me, I feel that I do not really deserve the praise. (Modified item)

I become panicky when I think of something I have done wrong (or might do wrong in the future). (Original Self-concept item)

> I become scared when I think of something I have done wrong (or might do wrong in the future). (Modified item)

As can be seen, modified items were those perceived to require only a minor change in wording, the central thought of the message being retained. For example, in the first modified item above, the idiom "practice what I preach" was one which was perceived as not likely to be uniformly understood by Ethiopian youths and hence was replaced with the phrase "do what I tell others to do," preserving, it is assumed, the original meaning of the item. In the two items following, only a single word has been changed, these because it seemed that they would be less likely to be understood than would the words which replaced them—"praise" for "compliment" in one item, and "sacred" for "panicky" in the second.

TEST DESIGN: THE IMPACT OF LANGUAGE

Language is a crucial factor in designing successful tests for developing nations. Ethiopia is a country of several linguistic groups, of which Amharic is the official and English the second official language. In the upper elementary and secondary schools and in the university, English is the language of instruction and in most instances is the language used in test construction. The effect on test performance of testing in one language or another is not entirely clear. For example, how does the language of the test affect the test's validity and reliability? For the test constructor, the central problem is to devise the most accurate measure of the trait identified as important. In devising such measures, the language in which the test is phrased should have no influence on the candidate's score. While this situation would be ideal, present evidence indicates that this is not so. We will review the work of the Testing Center on the problem of language and its implications for personal-social and cognitive style testing.

In one study conducted by the University Testing Center, identical tests of geography and intelligence were constructed in Amharic and in English and administered to random samples of 474 candidates for a government clerical position. The results were indeed revealing. Mean scores on the geography test for Amharic and English forms were virtually identical with, respectively, average scores of 20.84 and 20.88. But the case was entirely different for the test of intelligence. On this measure higher mean scores were registered for the test constructed in Amharic where the mean score was 29.47, as contrasted with the mean score of 21.82 for the group tested in English. These results were indeed unexpected and bear discussion even though any explanation offered must be ad hoc. One possible explanation resides in the nature of the task. The geography test was relatively straightforward and required little language sophistication. The ability test, on the other hand, required active manipulation of language, and it seems that the respondents had greater skill in manipulating Amharic than in manipulating English.

In the light of the above findings the decision to construct a test in one language or another seems to depend on the use to which the test results are to be put. If one is interested in measuring intellectual ability, that language most likely to unequivocally reflect it should be chosen. On the basis of present limited data, the Amharic Test should be chosen. On the other hand, there may be instances in which knowledge of English is important, and hence a test constructed in English should be chosen.

The above "rules of thumb" are merely suggestive, since additional questions need to be asked before constructing a test in one language or another. One question relates to test reliability. In studies at the Testing Center, higher reliability is generally achieved by tests constructed in English. However, this finding must be reviewed tentatively since the tests were first constructed in English, then translated into Amharic. It is possible that the low reliability of the Amharic tests can be explained by the fact that the fit between the languages is less than perfect in many instances.

A second consideration relates to test validity, or how well the test predicts some criterion. Whether a test is constructed in English or in Amharic should be incidental. Limited data on concurrent validity was revealing. Identical ability tests constructed in English and in Amharic were both correlated with certain other measures of Amharic or English to determine how closely scores on the different measures would be related. The data of this analysis are summarized in Table 1.

The one clear generalization emerging from the data of Table 1 is that tests constructed in the same language correlate more highly than tests in different languages, even though the tests were ostensibly identical. One implication of these findings is that tests constructed using different languages should not be combined to obtain a total score, since there is some unknown portion of the correlation which is due to language differences.

A third consideration in determining whether a test should be constructed in English or in Amharic relates to how the individual will be ranked on the respective tests. If language bears no relationship to test performance, then a score should be the same, whether in English or in Amharic. However, additional research on this problem should be undertaken.

A fourth consideration relates to preference for taking a test in one language or another. In one study at the Testing Center a strong preference was found for one's own language—in most instances, irrespective of the test's purpose. Seventy-four university students, mostly freshmen or sophomores, were give a 20-item "Language Questionnaire" with these instructions:

Below are listed a number of tests which could be constructed in a) Amharic, b) English, c) Galligna, d) Tigrigna, or e) another lan-

Table 1. Concurrent Validity of Tests Constructed in English and Amharic.

Test	Reliability (KR-20)	Correlation With Amharic Ability	Correlation With English Ability
Geography (Amharic	.645	.450	.400
Geography (English)	.771	.300	.400
DAT-Verbal	.567	.315	.570
DAT-Numerical	.731	.353	.586
Amharic Reading	.821	.442	.431

> guage (if none of the above is your first or preferred language). For each of the activities below mark the language in which you would prefer to be tested in order to present the most accurate picture of your ability, knowledge, attitudes, aptitude, achievement, or personality.

The subjects rated their language preferences in the following types of assessment: 1) admission to HSIU, 2) admission to a U.S. graduate school, 3) personality, 4) musical aptitude, 5) aptitude for a position in the Ethiopian Government, 6) achievement at HSIU, 7) admission to a German graduate school, 8) interests, 9) intelligence, 10) a post in the major field of study, 11) vocabulary, 12) mathematics achievement, 13) science achievement, 14) knowledge of Ethiopian history, 15) admission to undergraduate study in France, 16) admission to undergraduate study in the U.S., 17) ability to get along with people, 18) attitudes toward some person, 19) attitudes toward some institution (such as the church or the university, and 20) attitudes toward some object (such as automobiles or airplanes).

Of the 20 items listed above, those related to assessment of personality, interests, ability to get along with people, and attitudes towards other individuals, institutions, and objects are nonacademic. The results revealed overwhelmingly that the majority of respondents preferred to be tested in the language of their area. For the six nonacademic purposes listed above, 66 percent of those whose first language was Amharic checked Amharic as the language they preferred for testing and 83 percent of those listing Tigrigna as their first language preferred Tigrigna.

For the construction of nonacademic tests in Ethiopia, the clear suggestion from these data is that—from the student perspective at least—tests in the students' first language are preferred. Since feelings, attitudes, and emotions are most likely to be expressed in the language of the respondent's area rather than in English or some other language, and since there are unique phrases, expressions, and idioms that lose meaning when translated, personal-social and cognitive style tests are likely to have greater predictive utility if constructed in native tongues. Whether results of tests so constructed are more valid and reliable than those in English or some other language remains unanswered.

4

Teacher Grades and Teacher Tests

Measures of educational achievement more limited in scope than those previously discussed are teacher grades and teacher prepared tests. Standardized or other externally-constructed examinations are usually not suitable for measuring achievement of the unit of study of a particular classroom. The trend in countries relying heavily on teacher grades (the United States, for example) has been to supplement them with standardized, external-type tests. In countries using external examinations almost exclusively there has been some movement toward internal assessment of learning (as in France).

Course grades are widely used as the means for reporting pupil progress and determining promotion, graduation and honors. A grade or school "mark" is usually a composite of scores on teacher-made tests and/or classroom contribution, quality of laboratory work, homework assignments and other projects. Thus it sums up the teacher's evidence of pupil performance in a course for a limited period. Many primary schools do not assign a single mark for a course; instead, descriptions are given of various behaviors and performances toward different learning goals.

TRENDS IN THE DEVELOPING WORLD

A number of developing countries have been taking a hard look at the examination policy used in their schools. Some observers have been critical of overemphasis on passing examinations at different levels.

> The single most important factor affecting the direction of education in E. Africa today is the realization that the vast majority of the children who go through the system will never see the inside of a secondary school, and that they will have to make a living on the land after primary school. An education system dominated by examinations and aimed at preparing primary school children for secondary school and secondary school pupils for university, cannot meet the needs of the majority of children (**41**, p. 202).

There are a number of different systems of examinations and requirements for school-leaving certificates. In a number of countries there seems to be a trend towards a balance between the use of internal and external examinations and between formal examinations and school records based on daily work as measures of achievement.

In Iraq (as in Bulgaria and many school systems in the United States), students who are high achievers in regular school work are exempt from formal examinations. Cumulative school record is a basic consideration.

Countries with a restrictive policy of admissions to secondary schools usually require an elementary school certificate based on an external examination. But in a number of cases school records provide supplementary evidence (Lebanon and Libya are examples). For admission to secondary school in Mali an elementary school-leaving certificate (based on an examination) is required; final selection is made by a national guidance committee after considering candidate's age, scholastic record, and aptitude test results.

In Malaysia a system of automatic promotion has been adopted and pupil progress is assessed internally by regular tests. Standardized tests are now being constructed within the Ministry of Education. India also has attempted to strengthen internal school assessment to correct over emphasis on external examinations. In the mid-1960s a UNESCO study suggested that an unnatural emphasis on examinations in Asia was an obstacle to improving educational quality.

An effort has since been made in many countries to improve examination procedures. While this does not mean that most countries have abandoned external examinations and have emphasized internal or teacher examinations, it does indicate that there has been emphasis on using information from examinations to provide guidance for education teachers, pupils, officials, and for school administrators and parents.

TEACHER TESTS AND THEIR USES

Measurement and evaluation are important for policy, but they are also an essential part of instruction at the classroom level. In the learning process it is important to measure learning throughout the course of study as well as at the end of a period of instruction. Teacher examinations are a means of periodic measurement and have a number of functions:

- To help the teacher evaluate the adequacy of his/her instructional techniques. Just as the fundamental purpose of educational systems evaluation is to provide information for decision-making on a global scale, so teacher-made examinations are useful for decision-making at the classroom level. Low scores on a test may inform the teacher that the unit was poorly planned, not effectively presented, or was introduced to unprepared students. Failure of learning may be due to ineffective teaching or to lack of student effort or both.
- To assess student progress. Information on individual performance over a period of time may be used by student and teacher to assess progress.
- To motivate and direct learning in the short range. Periodic tests may motivate a student to study and help set goals for future learning. A

test covering the objectives of a unit of study helps define those objectives for both student and teacher.

- As a basis for course grade.

COMPREHENSIVE MEASURES OF PERFORMANCE

A teacher's tests do not cover educational objectives such as use of library, skills in committee work, ability to express ideas orally. Nor do they usually measure curiosity, enjoyment of art; literacy style, or creativeness.

Some might argue that tests are hardly adequate for determining whether learning has been attained. The artificiality of test conditions may suggest skills or knowledge that may or may not be evident in behavior; and behavior may point to knowledge not indicated by test results.

Scores on teacher's examinations are supplemented, therefore, by other evaluations of student performance. For example, direct observation can be made and recorded of library use, cooperation with others in committee work, public speaking. Also more systematic appraisal of many products of course work such as term papers, essays, outputs of home economics courses and woodworking or mechanical shops and art work can be made.

It is difficult for the observer to remain objective, to select significant and objective criteria, and to include a broad sampling of behavior or performance. But the subjectivity may be reduced if the teacher makes an effort to determine in advance what to observe and how to observe it. Methods used in observation or appraisal include:

- ranking—define specific elements to be observed, rank by degree or quality or how well criteria are met;
- checklist—itemize specific features to be observed; then by yes or no or by a checkmark indicate whether action took place and whether it was satisfactory; and
- rating scale—observation or impressions over a time period based on perhaps a 5-point number or descriptive word scale (superior, excellent, average, fair, poor). Weaknesses to be avoided include the "halo" effect (tendency to generalize from one or two attributes) and the tendency to avoid extremes and rate student performance in the middle.

Anecdotal records provide an informal way of noting spontaneous significant or unusual behavior that may contribute background for evaluating achievement. Behavior is observed, not "judged" or rated. Other indications of changes in student behavior may include:

- attendance records;
- completion of assignments on time; extra assignments;

- library use: possession of library card; number of books checked out; number reported read (either required or voluntary);
- case histories; autobiographical data;
- extracurricular activities; and
- awards.

DIFFERENCES BETWEEN TEACHER TESTS AND STANDARDIZED ACHIEVEMENT TESTS

Teacher tests and standardized achievement tests usually serve different needs. The teacher prepares his/her own tests as part of the teaching-learning process for a particular class and to evaluate the achievement of pupils in that class. Standardized achievement tests are given to many students and are useful for comparing schools with each other or with national norms. They may be inappropriate for the teacher's evaluation of student achievement; they may not measure the particular class objectives. A teacher's test is tailored to fit specific needs and is based on the subject content and objectives of that class; a standardized test is based on subject content and objectives common to many schools. Teacher tests are flexible, suited for frequent use on specific topics (although they may cover comprehensive subject matter); a standardized test usually deals with a broad part of knowledge or skill.

If teacher tests use the same criteria as standardized tests, that is, if they are valid, reliable, objective, and easily administered, they are useful indicators of student achievement. But too often the tests may be hastily or poorly planned, may neglect course fundamentals and may sample only minute details. In addition teachers may rely on essay tests which are easier to prepare but may fail to measure adequately student progress. Frequent tests or too-heavy emphasis on "passing the test" may distort the learning objectives, turn the student against teaching (and learning), and generally defeat their purpose.

Teacher grades are often relative, based on comparative class scores. But grading is not necessarily norm-referenced. Criterion-referenced teacher tests and other criterion-based measures could eliminate grading "on the curve." This means that before grading the teacher should have explicit criteria established. We have also pointed out that the unreliability of grades as measures of achievement may be caused by ineffectual teaching, variability of marking standards of the teacher and/or school, lack of objectivity of teacher evaluation, inconsistency, and so forth.

Standardized tests measure levels of knowledge, skills and competence. They are empirically developed, their reliability and validity checked and norms developed. They may measure the same knowledge and skills as teachers' classroom tests, for they usually try to cover only what schools teach and generally indicate whether students have mastered some body of material covered. Norms for the content of these tests are based on typical or composite course materials for a class, so some tests may not cover special objectives of local schools and ability levels of local pupil populations. Yet some studies suggest that skills needed to score high on a standardized achievement test and the skills necessary to earn high grades in school are similar but not alike (Jencks, 39, 1972).

Teacher grades as a measure of achievement present a number of problems. Usually influenced by factors besides performance, they may reflect class conduct and effort as well as achievement and ability. Furthermore different standards are used in different schools or by different teachers in the same school. Therefore grades awarded by different schools seldom reflect comparable levels of achievement. In the developing nations differences between schools and between teachers and their grading could be very great.

The advantages and disadvantages of standardized tests and teacher tests are listed below:

Advantages of standardized tests:

- all participants are evaluated on the same set of criteria;
- tests are objective; and
- standardized tests used at intervals throughout the pupil's school career provide longitudinal information.

Disadvantages of standardized tests:

- achievement of measures provides limited information about a student and his skill knowledge;
- tests are given in an artificial situation;
- tests may control instruction and course content;
- may not cover certain objectives of local schools; and
- may not measure knowledge of students who have not studied the usual textbooks.

Advantages of teachers' grades and school records:

- teacher grades are based on cumulative record performance;
- they show typical behavior and performance in daily classroom situation;
- they provide teacher and student with periodic assessment of achievement; and
- they show more than skill knowledge—they may include assessment of behavior that cannot be measured by tests.

Disadvantages of teachers' grades and school records:

- changes in short-range cognitive and effective changes are recorded and observed by the teacher only during the term of school year, not over the school career; cumulative records may not be a substitute for longitudinal records built up by standardized tests;
- teacher grades often compare students within a class instead of providing data on how much has been learned;
- subjectivity or bias of teacher: grades may be more influenced by good conduct, docility, neatness than on how much a student knows; and
- pressure from parent or community may influence grades.

ETHIOPIAN EXPERIENCE

A further problem with the use of teacher grades as a measure of educational achievement is illustrated by problems noted concerning the practice of using school marks for student evaluation in Ethiopia.

For some time school marks in Ethiopia have been given some consideration in determining the passing score for Grade 6 candidates. National Examination results count for 70 percent and school results 30 percent of the score. However, the testing department of the Ministry of Education has objected to this combination of results, claiming that school results do not reflect actual student achievement. They suspected that the teachers in various schools inflated the school result, hoping to increase the number of passers from the given school and thereby causing an upset in the National Examination results (Ministry of Education Reports, 1970, 25, p. 29). But the argument presented against using school results does not appear to be based on sufficient evidence, as no information is available to verify if teachers have been consistently generous in grading. The judgment of the testing authorities appears in this case to be based on a single year's grades, which is insufficient evidence for suspecting the authenticity of the school results. On the contrary the teachers might be better judges of student achievement than a body of examiners who have little or no knowledge of the examinee's ability. One may conjecture that a better selection procedure may be devised if teachers and Ministry officials collaborate to prepare the instrument.

Meanwhile it may be helpful to consider the appraisal made both by high school and elementary school teachers of a candidate's achievement in an external examination. Incorporating the evaluations made by Ethiopian classroom teachers may give a better picture of the general abilities of the students.

POSSIBLE WAYS TO ADJUST GRADES

Teacher grades can supplement standardized tests but if grades are to constitute a useful measure, adjustments would be needed. Perhaps some set of indicators could be developed to measure school differences. Some factors to be considered in such an adjustment are:

1) school policies on grading;
2) rating schools on:
   a) number of students going into higher educational institutions;
   b) number of former students employed in various job categories (or not employed); and
   c) general behavior of school leavers:
      use of library—number who have library cards,
      subscription to newspapers—read newspapers,
      use of radio and TV,
      membership in professional groups or other

organizations,
voting record; and
3) teacher differences:
general qualifications,
credentials,
experience.

Teacher grades and teacher tests may be of limited usefulness as a measure of educational results in developing countries where educational opportunities may be sharply delineated by economic or cultural factors. A mix of balance between standardized or other external examinations and internal, or teacher testing, seems to be the trend in many countries both as a guide to educators and planners as well as a measurement of individual educational achievement.

5

Socioeconomic Status and Achievement

Numerous and varied studies have addressed the general problem of accounting for differences in the educational performance of individuals. We concentrate in this chapter* on those studies that seek explanations in the "background factors" which are defined here as only those characteristics of children, families, schools, teachers, and communities substantially beyond the control of the educational establishment. These include: the age, sex, and ethnic origins of the child, the education and occupations of his parents, family income, the "cultural" atmosphere of the home as represented by books, periodicals, supporting equipment, musical instruments, etc. These factors are indeed associated with school characteristics. For instance studies indicate that a poor rural population will score low on the various scalings for socioeconomic status (SES): low parental education attainment, low-wage occupations, low family income, and a general absence of books, along with large families.

STUDIES IN DEVELOPED COUNTRIES

Research studies in developed countries have demonstrated that achievement by students in their academic work is accounted for by these background factors. This remains true even after adding other factors to the analysis to better represent the different aspects of socioeconomic status or the cumulative effects of educational experiences on the student.

There is a general consensus among the majority of the studies, nevertheless. The same kinds of background factors emerge again and again to account for whatever portion of the variability among students (and among schools) that can be explained by associative models.

Measuring Socioeconomic Status

A bewildering variety of ways to measure SES confront the researcher:

If it is impractical to measure family income directly, which of the many surrogates could be used? If he takes the occupation of either or both parents to represent income, which classification system and scaling could be used?

---

* The material in this chapter was prepared by Ezra Glaser and draws, in part, from the report, "Associations between Educational Outcomes and Background Variables: A Review of Selected Literature," by Edward C. Bryant, Ezra Glaser, and Morris H. Hansen.

How is one to represent the cultural climate and intellectual stimulation of the household? Should any specialized cultural and artistic interests be sought, by asking about musical instruments, artist materials, or athletic equipment?

Beyond the choice of variables, what populations are to be probed? How to sample? How large a sample? What procedures for the gathering of data? What kinds of validations and quality controls?

Despite these difficulties almost every major study has included some measure of SES. From these studies one learns to expect SES alone to account for between 10 and 25 percent of the variation in academic scores.

Most common measures of SES include occupation of parents, education of parents, and items in the home; but race and sex are also important components of the "socio" part of socioeconomic status. Most studies treat race and sex separately, not because they are not vital social characteristics, but because their potential importance argues that they should be separately measured.

In general differences in achievement scores of boys and girls, men and women, are small. When specialized subjects are considered some differences might be large enough to mention. The most important of these is the advantage boys have over girls in science scores. It seems plausible that cultural influences are at work here; it will be interesting to observe whether these differences diminish with the changing attitudes concerning occupations suitable for women.

Even after the SES factors have been statistically controlled a significant increment of explained variance is associated with race. Race is highly correlated with measures of SES. But the principal studies have shown that there is an additional contribution to the explanation of variance beyond that which can be attributed to SES. Studies show that race contributed an additional five to ten percent to explained variation after SES and family structure and stability had been accounted for.

What is suggested here is that there are other mechanisms at work that detract from the performance of minority youth than are measured in the usual SES scalings.

A study in Great Britain reported finding strong relationships between reading comprehension scores and three parent-student variables: aspiration for the child, literacy in the home, and parental interest in school work and progress. (The percentages of explained variance within schools ranged from 12 for lower junior girls to 26 for top junior boys.)

Innate Ability and Achievement

Common sense would seem to insist that there are important differences in the innate ability of people, after allowance is made for any influences that might be attributed to their sex, race, age, home environment, school conditions and possible other factors. No way has been found to measure innate ability uncontaminated by acquired knowledge.

Project TALENT collected 13 measures of ability from a sample of 2,900 12th grade American males. They ranged from two information tests—practically entirely acquired knowledge—through a group of intermediate items, to visualization in two dimensions—the most abstract measure of individual ability. SES measures explained .17-.28% of the variation in the information tests; it decreases through the sequence to .06% for the least knowledge-oriented tests. Even the most abstract tests contain the risk that they are influenced by both learning and the practice of taking tests. Perhaps a substantial part of the unexplained variation in achievement after accounting for SES and similar factors can loosely be attributed to "individual differences" consisting largely of the dimensions of ability and motivation.

Earlier school performance has sometimes been used as a measure of innate ability. True, performance in sixth grade (test scores, grades, teacher appraisals) is a very good predictor of 7th grade performance. But the 6th grade performance already contains a mixture of ability, motivation, self-esteem, and other factors that represent an accumulation of knowledge, attitudes, adaptations, and preferences that could hardly be considered innate in any useful sense. Indeed the use of such predictors explains little except the likelihood that students tend to be consistent performers: good students (however assessed) are likely to be found to be good students next year.

Attitudes Towards Schooling

Besides innate ability such affective states as attitudes, motivations, self-perception, aspirations, intentions, and expectations may explain achievement variations. Various studies have investigated such relationships. Many of them have found significant relationships. For illustration we turn to a study by Mayeske. He used data from the Coleman study, "Equality of Education Opportunity," and analyzed eight student variables. Four attitudinal variables were used to attempt separate explanations of achievement scores. The variance explained (squares of simple correlation coefficients) are shown for grades 3, 6, 9, and 12: Expectations (3 to 15 percent), Attitudes toward life (1.7 to 22), Educational plans and desires (6 to 26), and Study habits (5 to 14). In the first three, the lowest association was that for the 3rd grade and the highest was for the 9th.

It would be reasonable to expect that these attitudinal factors would be related to both SES (home and community characteristics) and the school. Indeed Mayeske found correlations between the four variables listed above and SES, reaching to a high of .29% for educational plans and desires for the 9th grade. But these are not so high as to suggest that most of the relationship of attitudes to achievement scores really represents an indirect way of recognizing the effect of SES.

School Influences

Furthermore in an attempt to isolate the influences of school variables, Mayeske found that 40 percent of the student variability in SES scales was associated with the schools they attended. Mayeske also provides an approximate measure of the incremental contribution of school characteristics after SES has been allowed for. The combination of SES and a number of "family process" variables amounted to 48 percent of the variability in scores. The addition of school factors brought the total up to 54 percent, an increment of six percent of the total variability in achievement scores for individual students. These results explain more of the variability than most other major studies, perhaps because of the many refinements in his analysis and the substantial size of the sample. Other large studies generally found that school factors accounted for five percent or less after adjusting for individual background factors.

A more direct view of the school influences can be offered by basing the analysis on the differences between schools. The impact that schoolwide background variables can have on individual student outcome measures is limited approximately by the proportion of total variance of the outcome measures that is accounted for by the variance among school means of the outcome variable. In some illustrative cases approximately 10 to 35 percent of the variance of the individual student achievement scores were accounted for by the variance between school means. It is this 10-50-35 percent portion that schools can explain, even with the inherent relations with other background variables.

After School Influences

The Mayeske results presented above used educational achievement as the criterion variable—the variable to be explained. It is also useful to consider other outcomes of the educational process, and to inquire into the factors that explain their variability. A model can be constructed which relates the education and occupation of a subject with the education and occupation of his father, the number of siblings in the family, and all of these factors to the income of the subject.

It is hardly surprising that educational attainment (highest grade reached) is related to occupation, income, labor force participation, unemployment (negatively related), and rates of pay. For studies made of teenage and young adult populations, their attainment at the time of the survey is also correlated with their participation in later educational activity. It has also been established that favorable attitudes toward school and toward work are correlated with their participation in later educational activity. It has also been established that favorable attitudes toward school and toward work are correlated with SES as well.

The importance of these findings lies in their assurance that all of the relationships that have been investigated for the explanation of achievement scores are not isolated from the realities of later life. Indeed, the same kinds of background factors that explain success in school also promise favorable job experiences, continuation of education, and useful social attitudes.

THE EXTENSION OF FINDINGS TO DEVELOPING COUNTRIES

The analyses presented in the previous section do not address questions relating to the improvement of the educational system. Rather they focus largely upon the kinds of students that come to the schools. There is little point in suggesting that a developing country would be better off if the parents of the students were better educated, or if their home life contained more intellectual challenges and cultural opportunities, or if the income were higher.

What insights, then, come from the United States studies and how can they be used to raise the educational achievement and attainment of people in developing countries? Several answers suggest themselves:

The importance of differences in social and ethnic origins in the United States (and other economically advanced countries) warns that progress in countries with considerable socioeconomic heterogeneity will likely be uneven. Those presently better off in SES terms will be able to take better advantage of all advances in education with little delay and their gains will probably be substantial. Those at an SES disadvantage will probably not be as fortunate in realizing potential gains from an improved educational system. Consequently there will also be an increase in inequality of educational achievement and attainment between more favored and less favored groups. If any other result is to be obtained there will have to be a deliberate effort (presumably designed with considerable understanding of the sociocultural characteristics of the groups to be served) with special programs and additional resources directed to the less favored.

The United States studies also indicate that without truly heroic efforts, advancement of educational status of a country's population will be gradual at best. As each succeeding generation offers a better start for its children, they will have the potential to take advantage of the formal educational opportunities. Indeed it would be easy to show that a stationary and unchanging educational system would produce better results each year as the successive waves of entering children were better equipped to succeed.

Although this might generally be true, there have also been examples of very substantial gains in short periods. Very substantial gains have been made by the children of immigrants to the United States over their parents' educational status in a single generation, possible traceable to the radical changes in social climate and expectations from the "old country" to the U.S. This involved, of course, a shift from one culture to another.

Other rapid strides might come from the initiation of large-scale children and adult literacy programs. China is such an example: "In 1973

Chih Chun, deputy head of the Science and Education Group under the State Council, reported that there were 127 million children in primary schools and 36 million in middle schools, five or six times the number in primary schools before liberation and 22 times the number in middle schools." Also the social background of students in higher education is reported to have changed radically: "In 1958 peasants and workers are said to have comprised only 19.5 percent. At present, it is reported that more than 90 percent are of worker, peasant, or soldier origin" (Encyclopedia Britannica, 1974 Yearbook, 21, p. 185).

Another lesson from the United States studies is that factors outside of the educational system, notably economic, can evidently prepare the country for improved educational status. An improved standard-of-living itself would probably create new opportunities for children to learn both within the formal educational machinery and outside.

Cultural progress not obviously related to schools or formal education would also be expected to influence attitudes, motivations and opportunities that in turn would improve achievement.

Health factors, concealed in the SES measures for the United States, are well known. (Surveys in the past have repeatedly shown poor health conditions among the poorer and less educated groups.) Further, present knowledge of the damaging effects of low nutritional standards among children suggest that real gains in education can be achieved simply by better nutrition.

The above findings suggest how educational systems in developing nations might be improved, but in any event it is always hazardous to transfer the results of complex social studies from one country to another. Having offered some tentative ways of interpreting the United States data in the context of a developing country, we now spell out one of the hazards in doing so.

Although there are great differences in the social status of various groups in the United States, the social status itself is poorly defined. There is opportunity for upward mobility and there are many governmental programs to encourage it—free public education being a prime example. Moreover, although the extremes are far apart, a large fraction of the population occupies a middle income ground of SES, and the middle ground improves significantly with each succeeding generation. As noted above many urban children of immigrant parents have moved from quite low SES to economic, social, and cultural success in a single generation.

Many developing countries have much more rigid social classes, often with social, legal, or economic restrictions limiting upward progress.

Another consideration is that in some developing countries the population is divided into a "privileged" class and a disadvantaged class; the two groups are usually ethnically distinct. A graph of the SES of such a population reveals a U-shaped distribution, with many people at the extremes and few in the middle. The same is true of educational attainment and achievement. Statistically the pattern differs from that for economically advanced countries where the population is closer to the normal distribution of heavy concentration in the middle.

The variables that represent SES in developing countries tend to be more highly intercorrelated than in developed countries. The disadvantaged are typically poor in all respects: income, health, housing, education, possessions, occupations, life expectancy. There is almost no overlap with the distribution for the favored population, which scores high on all of these variables.

Given the high correlation between variables representing education and those representing SES, a fairly detailed and accurate picture of the educational status of the population could be developed from their SES. The absence of the middle-range values makes the distribution act like two loosely aggregated points; a line drawn through the two clusters will explain an inordinately large share of the variation of one (say, educational attainment) in terms of the other (say, SES).

As in the advanced countries the relations demonstrate only associations between factors, not causality. But in developing countries there is no need to delay planning social improvement programs to quibble over the "cause" of poor education in the immobilized lower social classes; in some areas schools do not even exist; and in others they are too costly for poor families.

The statistical properties noted above merely reflect the fact that special programs would be required to break through the existing patterns if the disadvantaged groups are to enjoy more advanced education. Indeed if there are large fractions of the population in groups far from the national average, one might expect little progress from the low SES groups. Children of poor, landless, illiterate, rural parents would not seem to be in a position to respond to gradual improvements in the established educational technology. Many countries have already proceeded by nonformal education and by upgrading the educational system directly, deliberately breaking the pre-existing relationships in the face of social and technological obstacles. The desired goal is a chain reaction in which the higher educational status of children, compared with their parents, would lead presently to higher social and economic status. Indeed it is probably difficult to move forward in one area without doing so in the others, whether or not the country's plans included such goals.

SOME INTERNATIONAL COMPARISONS*

The comparison of national averages of literacy rates, educational attainment, and SES measures such as per capita income and family size inevitably depart from the true relationships. Only if there were very small variations among the people within a country would the national comparison reveal the relationships between educational variables and SES of the individuals. We shall see that the well known heterogeneity within underdeveloped countries shows up even in a summary analysis.

---

* Further international comparisons are presented in Chapter 6.

We present examples of great differences between urban and rural populations and also an example of the consequences of ethnic differences. In a country with U-shaped distributions of education and SES, the averages tend to fall in a middle ground where few people are found. Some illustrative data will show that the SES-education relationships are sufficiently persistent to be revealed by crude analytical procedures.

The examples of countries selected below are governed largely by the availability of data from UNESCO sources, taking advantage of attempts to develop comparable data on social conditions for a number of countries. Kenya, for example, is the only nation for which a direct comparison could be made of African and non-African populations.

The UNESCO definition of literacy: ". . . ability to both read and write is used as a criterion of literacy; hence all semi-illiterates—persons who can read but not write—are included with illiterates. Persons whose literacy/illiteracy is undocumented are excluded from calculations; hence the percentage of illiteracy for a given country is based on the number of reported illiterates, divided by the total number of literates and illiterates." (UNESCO Yearbook 1970, **69**, p. 31). The standard form of the statistic, when it is available, is the percentage of the population 10 years of age and older who are illiterate.

A second measure of base-level education used by UNESCO is the percent of the population 25 years of age and older with less than four years of educational attainment. Where both of these statistics are available for a particular country, both are included in the tables of this section.

SES can be represented by per capita income (measured in U.S. dollars) and by average family size, also from UNESCO.

Although UNESCO and the other United Nations departments have labored diligently to develop comparable statistics, it is well known that they vary in quality from one country to another. There are components of the population that are difficult to measure accurately in most large underdeveloped countries. For this reason the tables presented below do not include countries where the relationship being illustrated rests upon relatively small differences in the reported data.

Table 2 contrasts some countries with relatively small portions of the population lacking four years of educational attainment or literacy (UNESCO definition), or both, with examples of countries at the other end of the scale.

Each of the better-educated countries reports family sizes averaging between three and four persons. Per capita income is reported only for Australia; it is high among the countries of the world.

Table 2. Some Illustrative Data from Selected Countries: Illiteracy, 4 years of Attainment, Per Capita Income, and Average Family Size.

Country (and year of latest education report)	Percent of Population 15 years and older with less than 4 years of schooling	Percent of Population 10 years of age and older who are illiterate	Average Family Income ($ U.S. and year)	Average Family Size (persons and year)
Some relatively high-performance examples:				
Australia '66	1.0		2970 '72	3.5 '66
Bulgaria '65	21.4	9.8		3.2 '65
Hungary '63		3.1		3.2 '63
Uruguay '63		9.6		3.8 '63
Some relatively low-performance examples:				
Algeria '66	91.8	81.2	290 '71	
El Salvador '61	83.3	51.2	290 '72	5.2 '71
Iraq '57	97.8		280 '69	6.0 '65
Kenya '62	90.1		810 '72	4.9 '62
Mali '60-61		97.8	71 '63	
Nepal '61	99.0	91.2	80 '70	
Rumania '66	81.8			3.2 '66

The situation is very different at the other end of the distribution, even restricting the comparison to national averages. The percentages of the population failing to complete four years of schooling among the selected nations is in the 80's and 90's, with two of them in the high 90's. Mali has a high-90's rate of illiteracy. Except for Romania the family sizes range about five to six persons. Except for Kenya the per capita incomes failed to reach $300.

Table 3 penetrates the curtain of the national average a bit by showing some examples of countries with large differences in education between the sexes. In Table 2 Algeria was reported as being a little over 80 percent illiterate. Now we can observe that this average conceals the rate of about 70 percent for men and a bit over 90 percent for women. Similarly Nepal's

Table 3. Some Examples of Countries with Large Differences in Illiteracy Between the Sexes.

Country (and year of latest report on education)		Percent of population 10 years of age and older who are illiterate	
		Males	Females
Algeria	'66	70.1	92.0
Bulgaria	'65	4.8	14.7
Cambodia	'62	30.1	87.3
Nepal	'61	83.3	98.5
Paraguay	'62	19.0	31.3

average was about 91 percent; it is composed of about 83 percent for men and over 98 percent for women. Even in a country of such low literacy, a large sex difference exists. Among the more favored countries Bulgaria was only 9.8 percent illiterate, but even here there was a large sex difference: under five percent for men and almost 15 percent for women. In a more middle range Cambodia provides the most extreme example. (These sex differences in literacy are not typical; Table 3 contains the most prominent differences.)

Table 4 calls attention to the educational differences between urban and rural cultures within the same country. The three countries for which these data are available in the Yearbook have illiteracy rates which might be called low, middle, and moderately high. In all three cases the urban literacy was far better than rural literacy. As noted earlier the classifications "rural" and "urban" probably stand for a variety of other socioeconomic differences as well: ethnic composition, income, family size, etc.; no causal mechanism is implied.

Table 4. Some Examples of Countries with Large Differences in Illiteracy Between Urban and Rural Populations.

Country (and year of latest report on education)	Percent of population 10 years of age and older who are illiterate	
	Urban	Rural
Bulgaria '65	5.2	13.8
Ecuador '62	11.9	44.5
El Salvador '61	28.8	66.3

Table 5 reports the only country with the standard educational measures available for major ethnic subdivisions of the nation. The non-African portion of Kenya is much better off—in basic educational attainment—than the African portion. The difference is not substantially modified if one observes the data for females.

Table 5. Kenya: Educational Attainment of African and Non-African Subpopulations (percent of population 15 years and older with less than 4 years of schooling, 1962).

	African	Non-African
Total population (males and females)	92.8	37.4
Females	97.6	46.0

Table 6 suggests a way of looking at illiteracy data to detect recent changes in the education of the population, even though the data are all reported for the same date. In 1962, a quarter of the population of 15 years of age and older were illiterate. But on the same date those who were over 15 and not yet 20 years of age had an illiteracy rate of only half of that. These recent gains made a little more difference to the women than to the men.

Table 6. Paraguay: Illiteracy of Selected Age Intervals (percent of population 15 years of age and older who are illiterate, 1962).

	15-19 years of age	15 years and older
Both sexes	13.1	25.5
Men	12.0	19.0
Women	14.4	31.3

In this final section of the chapter we have given some of the indications that the association between SES and educational achievement and attainment that was established in studies of the United States also apply to countries that range from the least developed to the most developed, in economic and social terms. There was also the suggestion that the entire country was not generally a sensitive unit for analysis; but major components of the population can be separately examined with substantial increases in our understanding of the educational situation.

6

Cross-National Evaluation: The IEA Study

In 1966 the International Association for the Evaluation of Educational Achievement (IEA) began a six-subject, cross-national survey of educational achievement. Some 20 nations, four of them less developed, were involved in this evaluation of achievement in science, literature, reading comprehension, civic education, English as a foreign language and French as a foreign language (Table 7). This chapter briefly reviews that IEA study. Its focus is the project's implications for measuring educational outcomes in developing countries as a tool for policy analysis.

The main aim of the project was to evaluate how different systems, teaching methods and curricula in participating countries contribute to learning. The primary objective was not to compare educational achievement but to relate input factors (such as the social background of students, teacher competence, curriculum characteristics and teaching practices) to outcomes (in terms of student achievement and attitudes).

Simultaneously the study was intended to facilitate future comparisons of different education systems. Previous comparisons had been confined to descriptive analyses of systems and methods within nations, but the IEA study sought to add empirical evidence of the outcomes or "products" of the systems.

According to Husen, the researchers hoped to look at different international practices and the variations in outcomes; then identify factors which caused those differences (Comber and Keeves, 12, 1973). The analysis would then center on factors associated with those differences.

Essentially the IEA studies inquired into national systems of education with the premise that certain factors can be used to predict and/or explain achievement. Three-pronged cross-national comparisons were made in three areas: student performance within countries, school differences within countries, and differences among countries.

The survey was a vast and complicated undertaking conducted in 14 languages and involving some 300 experts, about 258,000 students, 50,000 teachers, and 9,700 schools.

Three age groups were tested:

Population I	10 year olds
Population II	14 year olds

Population III	students in the last year of secondary school

Guidelines were drawn up for selecting schools and groups to be sampled. (Procedures were developed to assure that the samples would be representative and that standard errors could be estimated.) Then achievement tests were designed for six subjects, and questionnaires administered to students, teachers, schools, and national centers. By May 1973 analyses had been published for three of the six subject areas: science, literature, and reading comprehension.

All countries in the IEA project except Thailand participated in the

Table 7. Countries participating in IEA Project.[1]

	Science	Literature	Reading Comp.	Civic Ed	English as a For. Lang.	French as a For. Lang.
Australia	X					
Belgium (Fl)[2]	X	X	X			
Belgium (Fr)[2]	X	X	X		X	
*Chile	X	X	X		X	X
England	X	X	X			X
Federal Republic of Germany	X			X	X	
Finland	X	X	X	X	X	
France	X					
Hungary	X		X		X	X
*India	X		X			
*Iran	X	X	X	X		
Ireland				X		
Israel			X	X	X	
Italy	X	X	X		X	
Japan	X					
Netherlands	X		X	X	X	X
New Zealand	X	X	X	X		X
Scotland	X		X			X
Sweden	X	X	X	X	X	X
*Thailand	X				X	
United States	X	X	X	X		X

* Developing countries.

[1] Poland and Rumania supplied information about science teaching in their countries. Science tests were given in Poland but data were received too late to be included in the analysis.

[2] The Flemish-speaking and the French-speaking areas of Belgium are treated as two separate countries because of their dissimilarities.

science testing. Test items included earth sciences, biology, physics and chemistry, and were based as much as possible on common elements in the school curriculum of the participating countries.

Ten countries were included in the survey of literature. Tests dealt with both achievement and response to literary works. The major reason for including this subject stems from the fact that literature has a central position in the cultural life of a community. There were broad differences between nations in the way literature was approached and taught. There appeared to be no universal standard of achievement; instead, standards were relative to the aesthetic and cultural presuppositions of the nation.

Reading comprehension achievement was surveyed in 15 countries. The study focused on cognitive aspects of reading and did not attempt to assess aesthetic or affective aspects. It found that there was generally more similarity in objectives of teaching reading among countries than in other subjects.

Still in preparation are the analyses of civic education in nine countries, teaching English as a foreign language in ten countries, and teaching French as a foreign language in seven countries.

TESTS AND TEST CONSTRUCTION

One of the primary concerns of the IEA study was the development of measurement instruments. For some time there had been a need for tests that could be used to evaluate technical assistance programs in education in the developing countries—evaluation of student competence in subject areas instead of head counts of students and graduates.

In constructing its tests IEA took precautions against undue cultural bias. An international committee composed of subject matter specialists, teachers, test developers and curriculum specialists was set up for each subject area. Each participating country had a national research center and subject committees. The international committee for each subject was responsible for construction of the tests for its field. Individual countries analyzed curricula, proposed item material, and conducted trial runs of test items.

Science Testing

Items for the science tests were selected and grouped in terms of the content and behavioral objectives they measured: functional information, understanding, application, and higher intellectual processes, including analysis, synthesis and evaluation.

Phase one of science-testing consisted of draft test construction and pretesting. The International Science Committee sent rough drafts of items to national centers for comment. The tests were then revised and returned to centers for pretesting along with a manual of pretesting procedures and instructions dealing with possible difficulties, such as translation, use of popular and scientific terms and substitution of local plants, animals and materials for unfamiliar ones in draft test items. National centers could substitute local alternatives as long as they stayed within the general framework of the item and provided that they informed the International Science Committee and submitted translated copies of their revised items.

The next step was preparation by the national centers of an analysis of items and tests as a whole based on results from pretesting random samples of populations.

The final tests were then compiled on the basis of a) test analyses, b) subject area and behavioral objectives, c) availability of pencil and paper items, and d) need for anchor items common to tests for two or more populations. Again, national centers were allowed to change translation and presentation of items where deemed necessary as long as the International Science Committee was informed.

The International Science Committee also sought information on the influence of two recent developments in science teaching: first-hand experience and original investigation. To measure these developments, attitude and descriptive instruments had to be produced. These instruments had to take into account the varied social and educational traditions, as well as different degrees of economic and technological progress in the various nations.

Descriptive scales were used to obtain information from students about such school variables as methods of learning science (from textbooks or practical experience) and laboratory learning (whether structured or unstructured). Attitude scales sought information about interest in science and attitude toward science in school and in the world.

Development of instruments on attitudes toward science was similar to that for cognitive measures: from a large number of items elimination and revisions were made after preliminary testing in three countries and then further pretesting in other participating countries.

Both the descriptive and attitude scales were included in a science questionnaire completed by students taking science courses. It is interesting to note that an optional test of practical abilities which used simple apparatus was given to a sample of students in two countries. From data collected it appeared that these tests measured skills and abilities considerably different from those measured by tests without apparatus.

Literature Testing

Techniques were developed for literature testing to measure ability to read and to answer comprehensive and interpretive questions. These two measures of comprehension and interpretation were the cognitive measures. In addition, a questionnaire dealt with interest in literature and degree of transfer between what a student reads and his/her life—these were the affective achievement measures.

The analysis compared achievement scores between nations and correlated them with the availability of printed matter, with radio and television usage and with other behavioral characteristics.

A major part of the literature survey was descriptive—a study of student response to literary selections.* New methodologies were developed to identify basic elements of response; test instruments were then developed as categorizing devices were empirically tested. The approach was to relate student response to literary work and at the same time to relate student response both to socioeconomic patterns of the country and to the way literature is taught in schools.

Reading Comprehension Testing

There were three components of the reading test: comprehension, speed, and word knowledge. While it was easier to agree on objectives for tests in reading than in subject areas, language, translation, and text selection were more difficult.

For the comprehension test IEA decided that tested skills should be cognitive and not appraisal of style, feeling or literary technique. Test items consisted of reading passages selected for a) suitable range of difficulty for populations tested, b) number and diversity of items, and c) variety of content and treatment. An overall criterion was that passages be suitable to all participating countries and not peculiar to any one culture or country. While it was agreed that selected test items discriminated satisfactorily between good and poor readers and were particularly effective for 14-year-olds and science students, the test tasks were found to be too difficult for the developing countries of Chile, India, and Iran.

A number of passages of a low level of difficulty were selected for the speed tests to measure fluency in mechanics of reading. Trial runs in participating countries helped eliminate ambiguous or too-difficult passages. The final test consisted of a four-minute test of 40 items.

The word knowledge test was based on word pairs, either synonyms or antonyms. Each country submitted pairs, ranging from easy to difficult, in their English translation. Then from some 300 pairs the national committees indicated which pairs were impossible or difficult to translate into pairs of corresponding difficulty in the language of their countries. Items retained were given a trial test and an attempt was made to get a subset of items nearly similar in difficulty across all languages. But the "final result is one in which the equivalences from language to language is suspect" (Thorndike, 1973, 65, p. 33). Generally the range of difficulty was greater in English. Also, items tended to get easier in translation.

Percentages of correct answers in the final work knowledge test tended to be low in the developing countries. Overall the variation among countries was greater than for the reading comprehension test. The IEA reported that test items showed good and relatively consistent discrimination within countries but were not as reliable for cross-country (or cross-language) comparison.

---

* Response is defined as the interaction between the individual and the literary work, an interaction that may continue long after the individual has finished reading. The response consists of associations, feelings, and reflections that occur as one reads a literary selection. The response is usually not explicit but indices of the response can be obtained (Purves, 55, 1973).

Questionnaires on Background

Another essential part of the IEA effort was the acquisition of background data. As part of that effort IEA developed and used the National Case Study Questionnaire and separate questionnaires for students, teachers and school administrators (Thorndike, 65, 1973). These questionnaires provided information necessary to construct profiles of the participating countries. Passow, Noah and Eckstein (in press) have, in their report on the National Case Study Questionnaire, drawn up "national profiles" for the 19 countries which participated in the first stage of the six-subject survey. The size of the per capita GNP varied from about 1,400 to 4,300 U.S. dollars in the industralized countries, whereas it varied from 90 to 270 dollars in the LDC's in the study. The size of the non-primary sector as a percentage of the GNP was in most cases 90 to 95 percent in the richer countries as compared to 50 to 75 percent in the LDC's. The difference was even more marked if the size in terms of number of people employed in the primary and non-primary sectors respectively were measured.

In addition data on school systems, teacher characteristics, family, and societal variables were obtained and are now being related to achievement data in a search for patterns of similarity and differences among indicators and patterns of student achievements.

The background data are also being used in an attempt to explain differences within nations (urban-rural, large school-small school, single sex-coeducational, high-low attainment schools) and to determine if these differences of national educational performance are consistent across nations.

The National Case Study Questionnaire gathered pertinent information on: structure of school system, types of schools; processes of curricula change; pupil selection process; teacher training practices; instructional materials, patterns and nature of instruction; social and economic conditions (unemployment rates, elitist school, familial); urbanization and modernization; language and culture; and National Centers' judgments of education system. The school questionnaires sought information on: numbers of schools participating; number of students per school; student/teacher ratios; area served by the school (urban or rural); staff (numbers of part-time, sex, age); and nature of courses. Teachers were asked about: postsecondary schooling; age; hours of preparation per week; membership in professional associations; in-service training; and refresher courses. Information sought from students included: family background: parents' education, father's occupation, number of children in family; and out-of-school habits of an educational nature: hours spent reading for pleasure, watching TV, on homework.

The data were helpful in constructing a School Handicap Score (SHS)—a device used to adjust for socioeconomic gaps between countries and for the different communities from which the school drew its students. The following variables were selected to form the SHS: 1) father's occupation, 2) father's education, 3) mother's education, 4) use of dictionary at home, 5) number of books at home, and 6) family size. Thus regardless of the quality of the formal educational system in the LDCs, the IEA study was able, on the basis of the impact of the family background factors, to predict a large difference in mean achievement between LDCs and the more industrialized countries. Parents in the former type of countries may be illiterate and reading material is likely to be absent from the home. On the whole the verbal environment in which the children grow up is almost entirely oral and there are rather few occasions on which reading skills picked up at school can be reinforced by experiences at home.

Thus difference in achievement levels on a standardized test between students in developed and less developed countries could be expected, considering the overall socioeconomic setting for the school systems in the two categories of countries. The outcomes of the multivariate analyses indicate that the total effect of home background variables in both science and reading is greater than the total effect of all the school variables. Among the 10-year-olds 35 percent of the variation between students can be attributed to family background and 22 percent to school factors, including, of course, all the instructional factors. The corresponding figure for the 14-year-olds are 42 and 26 percent, respectively.

ASSESSMENT RESULTS AND PROBLEMS

Generally IEA findings reinforced the conclusion that schools have less to do with achievement than home and family, at least in developed countries. The study also substantiated other findings that sex differences are not large but boys do better than girls in mathematics and science tests and girls do better than boys in reading and literature.

"Opportunity to learn" (in the science study) was the school variable most clearly established as being related to achievement. Other school-connected variables that seemed to have some relationship to higher achievement levels included teacher training (science); controlled practical work rather than informal investigation (younger populations, science); a lot of homework (literature); and classes that use a textbook (literature). High retentivity (as opposed to selectivity) has little or no effect on the achievement of high-level students.

The comment by the authors of the analysis on the science tests—that the IEA study had been more fruitful in sociological than in pedagogical findings—holds generally for the other two completed analyses as well. Little information and scant evidence have been provided for evaluating the contributions of different school systems, teaching methods and curricula to learning achievement.

The reading comprehension test results provided "few clues" as to what it is about the school environment that results in better reading achievement of its pupils (Thorndike, 65, 1973, p. 9). The science tests results provide few guidelines for evaluating the relative efficiency of different teaching conditions and methods; indeed, the greater part of the schools' contribution toward science achievement appears to be the result of decisions about providing opportunities to learn and the degree of effort (homework, years of study) demanded of the student.

The limited results of the study point not only to the importance of factors outside the school in determining the achievement of students in a given school but also to the need for greater research efforts (62, p. 9). Limitations on methods and data are apparent. Analytical procedures may be so inadequate that the effects of home on learning are overstated (54). Furthermore without a longitudinal study of student learning it is difficult to assess school influences.

A number of unsolved problems appeared to have influenced test results. Science test questions, for example, were generally difficult for students in most countries. This may have been caused by constructing tests to fit a generalized curriculum rather than the curriculum of any one country.

Translation problems were not completely solved in the reading comprehension test—there was so much variability in the data as a whole that questions are raised about the comparability of tests translated into 15 languages. There was less consistency in test scores at higher age levels than lower, due partly perhaps to the more difficult ideas and more complex language of the texts. And translators may not have comprehended fully the ideas of the original text.

Cultural and national differences cannot be completely accommodated in cross-national tests. In the reading tests the analysis points out that cultural differences appeared in interpreting motivations and emotions and that climate and style of life could have been basically different from the English pattern from which most of the test passages came.

Ambiguities and inconsistencies that appeared in the study lead to questioning of whether the test instruments were adequate, whether achievements being measured were the real objective of education, and whether there are ways to measure contributions of nonformal and informal education. These and many other unanswered questions brought to light by the study may lead to new national and international assessment efforts.

As noted above the developing countries will find useful the IEA experiences in preparation of measurement instruments for countries with a wide variety of cultures, different stages of development and disparate school systems and educational objectives. The developing countries had significantly higher student/teacher ratios and shared common problems of rapidly expanding school systems and instructional methods. And while the IEA studies point to the need for additional work on developing instruments to measure educational outcomes for non-Western societies and developing countries, they also highlighted the very great problems involved in developing measurements of educational achievement in general and international measures of achievement in particular.

But unique problems arose concerning the four participating LDCs. However, the lack of linkage of the information as knowledge assessed to the functional requirements in the developing nations is a severe handicap that can only be overcome by testing instruments appropriate to developing nations. In the science analysis Comber and Keeves stated that the level of economic development and lack of a history of universal education showed up clearly in the test results (12, 1973). So great were the differences that cross-country test results of the four developing countries were compared, separately, with themselves, not with the 15 developed countries. A separate mean score was calculated for these four countries. Also, results from reading comprehension tests in the developing countries seem to indicate a relatively greater importance of school factors compared to home influences in the developing countries than in more developed ones.

THE TESTS AND THE DEVELOPING COUNTRIES

Up to now no representative comparative information with regard to student competence in the developing countries has been available. Chile, India, Iran and Thailand participated in the science survey and all but Thailand participated in reading comprehension. Chile took part in literature, English and French studies and Iran in civics. The IEA study was the first attempt at qualitative comparisons between developed and developing countries made according to an agreed-on international yardstick. Furthermore, the reading comprehension survey was the first comparative assessment of literacy levels among representative groups of students in developing countries. The non-industrialized or developing countries were consistently far behind the industrialized in average achievement over subject areas and levels of schooling (Table 8). In science the score of the less developed countries was roughly one standard deviation or more below the more developed. This means then that in science the average student in the developing countries scored between the 10th and 12th percentile lower than in a developed country. The difference was even more pronounced in reading comprehension, where only five to ten percent of the students in the developing countries scored at the level of the average student in a more developed country.

What explanations can be advanced for such a large difference? First, we must caution emphatically against any premature conclusions about the "productivity" or "efficacy" of the school systems in the two categories of countries on the basis of the mean scores presented in Table 8. The differences that we find between the industrialized countries are negligible in

Table 8. Mean Total Score and Standard Deviation in Science and Reading Comprehension Among 10-Year-Olds, 14-Year-Olds, and Pre-University Students in 19 Countries.*

	SCIENCE						READING COMPREHENSION					
	10-year-olds		14-year-olds		pre-university students		10-year-olds		14-year-olds		pre-university students	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Australia	—	—	24.6	13.4	24.7	10.7	—	—	—	—	—	—
Belgium (Flemish)	17.9	7.3	21.2	9.2	17.4	8.1	17.5	10.2	24.6	9.7	25.0	9.3
Belgium (French)	13.9	7.1	15.4	8.8	15.3	7.9	17.9	9.3	27.2	8.7	27.6	9.2
England	15.7	8.5	21.3	14.1	23.1	11.5	18.5	11.6	25.3	11.9	33.6	9.0
Fed. Rep. Germ.	14.9	7.4	23.7	11.5	26.9	8.9	—	—	—	—	—	—
Finland	17.5	8.2	20.5	10.6	19.8	9.8	19.4	10.8	27.1	10.9	30.0	7.5
France	—	—	—	—	18.3	8.7	—	—	—	—	—	—
Hungary	16.7	8.0	29.1	12.7	23.0	9.0	14.0	9.8	25.5	9.9	23.8	8.9
Israel	—	—	—	—	—	—	13.8	11.0	22.6	12.8	25.2	10.8
Italy	16.5	8.6	18.5	10.2	15.9	8.8	19.9		27.9	9.3	23.9	10.2

Japan	21.7	7.7	31.2	14.8	—	—	19.9	8.8	—	—	—	—
Netherlands	15.3	7.6	17.8	10.00	23.3	11.1	—	—	25.2	10.2	31.2	7.0
New Zealand	—	—	24.2	12.9	29.0	11.6	17.7	9.5	29.3	11.0	35.4	8.1
Scotland	14.0	8.4	21.4	14.2	23.1	12.1	—	—	27.0	11.5	34.4	8.2
Sweden	18.3	7.3	21.7	11.7	19.2	10.2	18.4	11.1	25.6	10.8	26.8	9.3
United States	17.7	9.3	21.6	11.6	13.7	9.5	21.5	10.5	27.3	11.6	21.8	12.0
Industrialized Countries	16.7	7.9	22.3	11.8	20.9	9.9	16.8	11.6				
Chile	9.1	8.6	9.2	8.9	8.8	6.0	9.1	9.3	14.1	11.1	16.0	8.8
India	8.5	8.3	7.6	9.0	6.0	6.0	8.5	9.4	5.2	7.2	3.5	5.8
Iran	4.1	5.4	7.8	6.1	10.2	5.6	3.7	6.9	6.7	4.4	4.4	6.0
Thailand[1]	9.9	6.5	15.6	8.1	12.4	6.1	—	—	—	—	—	—

---

* Table prepared by Torsten Husen.

[1] Thailand did not test national sample but sampled scores in the Bangkok area.

comparison with the gap between the two categories of countries. There is, however, no reason to believe that the more developed countries are all on the same level of "efficacy" with regard to their school systems.

A plausible explanation is that the tests are not doing justice to the children in the developing countries. The tests might draw upon knowledge and learning experiences more prevalent in the rich countries. Furthermore the test situation and the format of assessing the outcomes of learning might imply a certain cultural bias against students in developing countries. A number of IEA researchers feel that while they cannot entirely refute such hypotheses, the empirical evidence does not give much support. First, the content of the tests, that is, the individual test items, went through a long procedure of scrutiny and try-out before they were "passed" by all the national subject area committees and included in the international tests. Second, the rank order of difficulties of items tended to be highly correlated over countries, which indicates that differences in total scores between countries are not so much accounted for by differences in particular sub-areas or topics of a particular subject as by systematic differences in level of competence. The teachers were asked to rate, on a four-point scale, each item in the tests with regard to what opportunity the students in his or her class had had to learn the subject matter that was assessed by the item. As far as science is concerned, the average opportunity tended to be somewhat lower for 14-year-old and science students in the developing countries (see **12**, 1973). But these differences in opportunity can by no means explain more than a small portion of the difference in mean performance.

Other researchers and analysts are not so sure that there is no bias. Thorndike, for example, indicates that some test items used during trial runs were shown to be too difficult for students in developing countries but were still retained. Platt feels that much research is needed to develop instruments for measuring educational outcomes in ways appropriate to the needs of non-Western and developing countries. IEA's achievement tests were not a satisfactory fit for the students in the developing countries participating (**54**, 1973). Platt (and others) point out a number of problems:

- the tests were developed and edited in English and then translated into the language of participating countries;
- tests may have depended too much on reading ability and, based on the high rates of error on the reading comprehension tests in Chile, India and Iran, students in those developing countries may not have been able to read items in the science tests;
- the concepts and values sampled in the tests were not oriented to non-Western countries;
- North American and West European education orientation dominated test construction; and

- developing countries have had less experience with the multiple-choice format used.

It would be useful when evaluating educational outcomes in an international context, particularly when non-Western nations are involved, to look more deeply into the role of social and political patterns of behavior. Then achievement or competence in various subject areas could be interpreted in relation to social environment (including sex and class distinctions and family structure); to political structure and attitudes; and to education-related services (school lunches, health care and medical services, and so forth).

OTHER PRODUCTS OF THE IEA STUDY

In the final analysis among the most useful outputs of the IEA study are the processes and procedures developed and the data acquired. The international cooperative machinery and processes provide a base for continued and improved international assessments. And new research on, and re-evaluation of, the data may prove fruitful.

For all countries and the developing countries in particular a number of strategies and techniques used in the IEA survey could be utilized routinely. These include:

1. methods of analyzing national curricula by the goals to be achieved;
2. techniques for constructing instruments to measure progress toward achieving those goals;
3. procedures for drawing probability samples from target populations under consideration;
4. experiences in data processing that are appropriate for nationwide evaluation surveys;
5. routines for data collection in the schools, including experience in developing, distributing, and processing questionnaires; and
6. testing manuals developed for the preliminary work to be used as guides for countries seeking to improve their testing.

Furthermore, valuable data have been acquired; much of it has yet to be analyzed. For instance as part of the mathematics study a data bank was developed. Described as a technical breakthrough, it permitted data to be packed on a single tape. With the larger six-subject survey also to be packed in an equivalently small space, it will be possible to locate the bank at several places in the world. When other data are added to the IEA data bank, researchers will be invited to evaluate methods and approaches to teaching and to reanalyze and reinterpret the data.

AN ETHIOPIAN PERSPECTIVE ON THE IEA STUDY

Validity of the Findings. Despite the many contributions of the IEA studies as models for research in the study of educational input-output relationships, adoption of the IEA's methodology and findings in developing countries must proceed with caution. First, for many developing countries measures of educational achievement already have been constructed and used on a national level. Consequently these indigenous measures are likely to be more useful than those developed for wider and more general application (that is, the IEA measures of educational achievement). Second, as already noted greater variability in school resources and inputs will be found in developing countries, thereby increasing the likelihood that significant relationships between inputs and outputs will be found.

Indigenous Measures. Achievement tests used in the IEA studies included measures of competence in reading, literature, French as a foreign language, civics education, English as a foreign language, and science. These criterion tests represent the outcome of much work and conceptualizing, and the effort is to be applauded. However, in many developing countries nationally developed and administered tests of achievement are available for certain student populations. In Ethiopia, for example, achievement results are available for all students completing grades 6, 8 and 12. These tests of achievement in Amharic (the official Ethiopian language), mathematics, English, the sciences, and the social sciences are based upon subject matter prescribed by the Ministry of Education and hence presumed to be specific to Ethiopia. While the IEA study can yield useful comparative information and also provide support (or refutation) of provocative hypotheses about relationships between school inputs and outcomes, the extent to which information specifically useful to developing countries is provided is limited. Such information is most meaningfully provided by study of inputs and outcomes indigenous to schools and to specific developing countries.

Variability in School Inputs. Insofar as the generation of findings specifically useful in the allocation of resources and the modifications of curriculum practices is concerned, the productive yield of the IEA studies has not been great. In summarizing results of the literature study, Purves (55) wrote:

> It is apparent . . . that the efficacy of a school's practices must be seen in the light of the context of the school and the national school system. One must further conclude that the survey technique has not uncovered those school or instructional variables which best account for the superior achievement of the greater part of the students in school (p. 22).

Similarly, in summarizing the results of the reading comprehension study, Thorndike (65) writes:

> It must be confessed that the results of the study provide little guidance for the improvement of the educational enterprise. They point out the very decided importance of the input into any school system in determining its outcomes, but as in the massive study of schools in the U.S. included in the Coleman report, 1966, they do little to accentuate differences between schools in their effects on students. This is not to say that schooling is unim-

> portant. It may be merely that schooling is relatively standardized, so that extreme variations in quality do not occur. The results are consistent with the view that more extreme variations occur in home and familial backgrounds, and that the school is an important aspect of society that provides more nearly standard experiences and opportunities.

Such an interpretation may indeed be correct, particularly for many of the developed countries included in the survey. However, as noted, it is at variance with findings for certain of the developing countries in which school factors may assume greater importance in the prediction of school performance, in personnel, and in physical resources. As examples, consider these facts. On the Ethiopian school leaving examination, completed by all secondary school leavers, the number of passes (that is, "C" grades in five subjects) for students enrolled in individual schools will range between 0 and 85 percent. Great variation in the educational background of teachers will be found as well. Among junior secondary (grade 7-9) teachers employed during 1971-72, one percent had only an elementary education, eight percent had highest attainment in the high school grade and only one percent were college graduates, approximately 47 percent were graduates of teacher training institutes, having completed 10 years of general education and two years of teacher training, about 34 percent had some college training but did not graduate. At the senior secondary level, (grades 10-12) *three* teachers had only elementary school training and six percent had only a high school education. About 20 percent held a bachelor's degree, and about 10 percent were graduates of teacher training institutes. Fifty-one percent had some college training, but did not graduate.

Pupil-teacher ratios were varied as well. During 1971-72 they varied from 47.1 for the primary grades to 27.1 for the senior secondary grades. The junior secondary ratio was 32.1. There were, of course, significant variations in ratios from one school to the next. Such variation in pupil-teacher ratios, and the wide variation in teacher background (as well as other educationally relevant variables) provided data that make possible the testing of relationships hypothesized to exist between school inputs and outcomes. Such studies would not be possible in countries where greater uniformity is found for the input and output measures.

From the perspective of research, the variability in inputs and outputs found in schools in developing countries appears to yield some advantage. It should however be noted that while there is greater variability in school resources and inputs in developing countries than in many of the developed countries, the schools have also introduced some uniformity into the educational system. For example, there is standardization of socialization experiences and the exposure to common ideas. In other words while differences in school inputs indeed do exist, there are many commonalities as well, and for Ethiopia—and undoubtedly for other developing countries—the common experiences fostered in the schools are important.

Other Variables. While the results of various multivariate analyses indicate relatively greater contributions of home than school factors to the prediction of achievement outcomes, data for certain developing countries revealed a greater importance of school factors. The question remaining to be addressed concerns specific school factors that might be predictive of student achievement in the developing countries. In a country such as Ethiopia there are variables which are unique to the country and its school system. These include, for example, wide variation in staffing of schools and in the percentage of expatriate teachers, proximity to major city, degree of urbanization, and so forth. Motivational factors undoubtedly also are influential. Variables such as achievement motivation, self-concept, locus of control, and status concern (that is, desire for upward mobility) probably are critical as predictors and consequences of school outcomes. Therefore in studying outcomes in the developing countries it is important that the influence of these variables be determined as well.

7

Modernization: The Social and Political Impact of Education

The role that education is expected to play in social and political change is often indicated in educational objectives of developing countries. The educational objectives recommended by the Education Sector Review Commission of Ethiopia, for instance, are mainly concerned with social and political factors. Out of 13 qualitative educational objectives listed by the Commission, eight of them pertain to attitudinal, social and political factors (**23**, Chapter III). A major recommendation of the Commission on Primary Education for the Conference on Education and Scientific and Technical Training in Nairobi (**13**) was that African governments should "insure that the primary education system . . . (b) contribute to the strengthening of national unity . . . (c) bring about the social and cultural integration of children in the community; and (d) act as factors of change and of economic and social development."

Ideally of course indicators of educational outcomes are derived from objectives. Objectives such as those mentioned above do not, however, specify the expected outcomes in sufficient detail. Furthermore educators and researchers have so far done little by way of exploring the manner in which education might operate to influence social and political change, and much less by way of developing methods for the measurement of the outcomes.

Therefore this chapter begins with a description of social and political characteristics generally associated with developing countries, followed by an exploration of the conditions under which education can effectively serve as an agent of social and political change in such countries. Outcome indicators on which measurement efforts might focus are then specified. The identification of indicators and the discussions about approaches of measurement draw on several pertinent studies. It is found in particular that indicators associated with "modernization" appear to be characteristically those attributable to education. Measurement of factors affecting modernity also present possibilities for adaptation and use for the measurement of social and political outcomes of education.

SOCIAL AND POLITICAL FACTORS ASSOCIATED WITH DEVELOPMENT

Many of the major problems of development in developing countries result from social-psychological traditions with ingrained attitudes and value systems which are antithetical to social and economic development. National leaders look to education as a prime force in the complex task of dismantling regressive customs and encouraging new mores, while at the same time maintaining and strengthening vital cultural and social traditions.

Some typical traditional social characteristics of developing countries with affective implications are rigid social class stratification, atomization of social units, fatalism, depressed status of women, and alienation resulting from colonization. Examples of these characteristics are given below.

Economically and racially defined social class systems, such as that of the Indians in some South American countries, may result in occupational channeling and impede communication and cultural integration. Primary loyalties to kinship, caste, or tribe may compete with and tend to undermine national efforts toward unification and development. Urbanization in Middle Africa is one force for "detribalization" and acculturation, but it can have negative consequences as well—alienation and loss of a sense of identity. In South Asia the sense of national unity is obscured by the orientation of people toward family and caste. Kinship orientation reduces mobility which, along with fatalistic value systems, tends to sharply limit individual striving and entrepreneurial endeavors. Furthermore lack of faith and trust in others outside the immediate family or tribe lowers the level of interpersonal cooperation in government and business.

The religious and philosophical values of renunciation of worldly ambition and of passive acceptance of whatever life brings (common in some South Asian countries) may hinder social and technological change. Strictly delineated roles for women, such as Pakistani purdah requiring seclusion and nonparticipation in civic affairs, reduces the developmental and democratic potential of a society.

Long periods of colonization in many parts of the world, particularly Africa, tended to fragment and denigrate existing cultural values and patterns of life. People in urban areas may have adopted European values (for example, in their educational aspirations) which may not always be rewarding or functional for them as individuals.

THE ROLE OF THE INDIVIDUAL AND INSTITUTIONS IN MODERNIZATION

Two different views have been advanced regarding the process of modernization. One view holds that the key to modernization lies in personality changes—changes in individuals, values and attitudes; and the other maintains that changes in institutional structures must come first to provide opportunity and incentive for changes in individuals. Those who support the first view note that institutional changes do not bring rapid development automatically—that attitudes, values and motivation are more basic. Those who maintain the second view argue that institutional and other societal barriers retard development; remove those barriers and general development will follow.

No social scientist denies that the individual is a factor in social and economic change however. "For in the end, it is men, and in particular their deepest concerns, that shape history" (McClelland, **45**, p. 437). But many differ on whether individuals or institutions change first in the modernization process.

Modernization is, however, a complicated process involving changes in the entire pattern of national life. The developing nations with their great differences in history, culture, and resources are undergoing varied experiences in development and modernization. In one country very few people may be motivated towards development, and the governing elite may have to lead the way and promote attitudinal change and development. In another country, with large numbers of individuals favoring social and political change and economic growth, institutional stalemates may prevent action.

THE ROLE OF EDUCATION IN SOCIAL AND POLITICAL CHANGE

Traditionally it is held that education performs the dual function of acting as an agent for the transmission of culture and as an instrument of change. Some maintain that schools should simply reflect society as it is. Others argue the opposite, holding that education should actively take part in directing social change and in creating a new social order. This issue of conflict does not generally preoccupy educational policymakers in developing countries. The faith that is placed in education as an agent of change is of such far-reaching importance in fact that much greater attention has centered on the need to develop and adapt education to meet the needs of rapid development.

> "Education is the key that unlocks the door to modernization." Statements like this recent one are gaining acceptance as truisms by many nation builders, policy planners, and scholars interested in the modernization process. Once regarded as an essentially conservative, culture-preserving, culture transmitting institution, the educational system now tends to be viewed as the master determinant of all aspects of change (Coleman, 11, p. 3).

Social and political outcomes of education are, however, much more complex and difficult to evaluate than may appear, and require yet a great deal of study and empirical foundation. The conditions under which education can serve effectively as an instrument of social and political change need to be clarified.

The phenomenal expansion of education in many developing countries has occurred amid the largely transitional social, political and economic structures that prevail. The worldwide call has been for the reformation of education in developing countries. The notion of adaptation of education to the needs of society is not to be confused with the traditional view of education as an agent for the transmission of culture. It must, rather, be understood in the context of a deliberate search for ways of bringing about change through education in traditional societies. And as indicated earlier education can be dysfunctional as a factor of social change if it is not vitally linked and adapted to the conditions and circumstances of the society in which it develops.

Further, if education is to play an effective and dynamic role in social change the curriculum and educational approach employed must consist of the inculcation of attitudes and abilities relevant to the solution of problems which face the community at large. Education for social change is education for the reorientation of attitudes and values rooted in traditional cultural patterns; thus, there is need for more than the acquisition of skills in literacy and vocation.

> The essential task of the educational planner is to root the educational system in a complex matrix of influences making at once for change and also for new kinds of stability-influence related more to appropriate use of trained men than to projecting how many need to be trained (Anderson, 2, p. 27).

THE NEED FOR BROADENING THE SCOPE OF EDUCATION

One general indicator of social development is the degree to which institutions develop and cope with specialized social functions. The family unit in traditional societies may perform most functions of socialization, including the training of children and youth. The introduction of formal systems of education is itself a criterion of social change. It means a transfer of a major part of the function of socialization to educational institutions. Modern education is regarded as a new phenomenon to established traditions in developing countries; to be effective as an instrument of rapid change it should involve all segments of society. A test of this achievement is the access provided to education—the numbers and percentage for whom education is provided. In the past formal education in developing countries resulted in the formation of new political elites; educational opportunity was extended to a very limited segment of society rather than on the broad basis of universal access.

Further, it must be recognized that the type and degree of socialization that is possible within the formal system of education varies and so, accordingly, do criteria. School curricula has been used as a criterion. In a study of political socialization in Indonesia, for instance, Fischer found that political content did not seem to affect students below secondary and university levels. On the university level factors such as student interpersonal relationships and associations appear to have greater influence on political socialization than formalized aspects of instruction (Fischer, 31, pp. 114-115).

Having explored the role that education is expected to play in developing countries, let us now identify some social and political outcome indicators and examine the possibilities of measurement.

INDICATORS OF SOCIAL AND POLITICAL OUTCOMES OF EDUCATION

Measurement of the effects of education on social and political change begins with a specification of objectives and indicators of the outcomes sought. Studies which have been undertaken in the attempt to understand the process of modernization have identified two types of indicators: those which pertain to changes in the attitudes and behaviors of individuals and those which relate to changes in broad patterns of social conditions and relationships. Studies which more directly link education with social and political change have also dealt with these categories of outcome indicators.

First, we discuss examples of attempts at measurement from the point of view of attitudinal and behavioral changes in individuals. We then consider those pertaining to the global patterns of social evolution.

Attitudinal and Behavioral Outcome Indicators and Measures

It has been indicated earlier that certain attitudes and beliefs which do not appear to be positively linked with development aims are predominant among the populace of developing countries. Conversely attitudes which are supportive of the development process and often associated with the dominant patterns of behavior in industrialized societies are not generally observed. The dominant attitude and behaviors characteristic of people in developing countries are usually rooted in cultural patterns. Certain attitudes similar to those in industrialized societies are acquired through learning.

An extensive questionnaire-interview study involving some 760 South Korean professors and more than 750 journalists sought, for instance, to relate modernization with the following indicators among others:

1) ability to accept new changes;
2) national consciousness;
3) cooperative spirit;
4) an enterprising spirit;
5) critical attitudes;
6) innovating spirit;
7) resistant attitudes;
8) professional knowledge;
9) rational judgment;
10) spirit of participation; and
11) spirit of subjectivity (Hong, 33, 1967).

These are examples of the types of outcomes upon which the influence of education needs to be investigated. Of particular interest here is th finding that the professors and journalists were predominantly change-oriented with tolerant, nontraditional attitudes toward social change and occupational preference.

Addressing the question of the social and political outcomes of education more directly, a preliminary listing of outcome indicators was attempted in Ethiopia by way of questionnaires administered to a restricted number of educators, policymakers, planners, and university students. Respondents were asked what they considered to be outcomes pertaining to the individual, as well as social and political change in general. The listing compiled included elements pertaining to changes in attitudes and behaviors as well as in broad social and political patterns. From the point of view of the former category of indicators the Ethiopian survey suggests that the influence of education on the following specific indicators be assessed.

1) Achievement
2) Attitude toward work
3) Industriousness
4) Initiative
5) Leadership
6) Professionalism
7) Rationality
8) Risk-taking
9) "Secularization"
10) Self-reliance
11) Tolerance

How can the influence of education on such outcomes be measured? The nature of the indicators suggests the use of tests principally of the non-cognitive type. Important among such measures are indicators of modernity.

Modernism-Traditional Scales

Inkeles and his colleagues (**36**, 1969) defined modernity by a complex of traits, stressing that the development of a nation requires the transformation of the individual. They developed a "characteristic profile" (adapted in part from Kahl) to show manifestations (indicators) of modernity among cultivators, craftsmen and industrial workers in six developing nations. This was incorporated into a "modernity test"—a questionnaire administered (individually, through an interviewer) to 6,000 young men in Argentina, Chile, India, Israel, Nigeria, and East Pakistan. The test covered attitudes, values and opinions on most major aspects of modern life (Inkeles, **36**, p. 215).

On the basis of his modernity test Inkeles lists the following as some tentative behavioral consequences of "modern" attitudes:

- joining two or more organizations
- voting often

- discussing politics with wife
- contacting officials about a public issue
- reading newspapers daily
- scoring high on geographic, political, consumer information scale.

Other measures of modernity have been suggested. These are: risk-taking, rigidity, trust, attitudes toward government, and work attitudes.

A risk-taking scale measures propensity to take risks as opposed to seeking security (Manaster and Havighurst, 47, 1972). An example of test questions is:

Which do you prefer?

A job where I am almost certain of my ability to perform well.

A job where I am usually pressed to the limit of my abilities.

Rigidity scales may provide indications of individual ability to accept change (developed by Williams in 1962, and presented in Manaster and Havighurst, 47, 1972). Items often included are given below with examples of attitudes toward change.

- *need for a stable, predictable environment:* I don't like to undertake any project unless I have a pretty good idea as to how it will turn out.
- *conservation and conventionality:* I would be uncomfortable in anything other than fairly conventional dress.
- *opposition to change of judgment:* It's a good idea to have a strong point of view about things because that makes it easier to decide what's right or wrong.
- *opposition to change of plans:* Few things are more upsetting than a sudden unexpected change of plans.
- *adapting to new and unusual situations:* I do not like to have to adapt myself to new and unusual situations.
- *learning new ways of doing things:* I dislike having to learn new ways of doing things.

Scales pertaining to faith in people may provide some indication of attitudes about cooperation with others. These scales seek responses to statements such as:

- Most people can be trusted/not trusted.
- Most people are more inclined to help/compete with others.
- Most people try to take advantage/try to be fair.

National involvement and political participation may be measured by scales tapping attitudes toward national symbols (for example, flag, national language) and by self-reports regarding such practices as voting and participation in public demonstrations.

A Case of Test Adaptation in Developing Countries

An attempt has been made in Ethiopia to adapt and administer noncognitive tests first developed in the United States to students as they complete the secondary level, along with the regular aptitude tests and the Ethiopian

School Leaving Certificate Examinations. Though the immediate purpose is to attempt to predict success in the university, the tests are so designed as to enable assessment of outcomes as they relate to the broader social and political goals of education and included measures of achievement, motivation, attitude toward social mobility, sense of control of the environment, self-esteem, reduction of fatalism and the life (see Chapter 3 for a description of the measures).

These measures, though not developed specifically for purposes of assessing educational outcomes, do reflect goals of education, which if not always made explicit exist nevertheless. For example, it is expected that students gain in self-esteem as a result of their school experience, that they experience a measure of control of their environment, and that they approach their life situation with optimism rather than fatalism.

Measures of fatalism, self-esteem, and locus of control—among many others—were administered to a random sample (approximately 20 percent of high school students completing the twelfth grade). At the same time students were taking an aptitude test along with tests of achievement in various subject fields which had been a part of their high school programs. Of all candidates taking the achievement tests during this period (including those taking the tests for the second and third time) only about 1 in 10 would be successful. Competition is thus very keen. Given such a situation it might be expected that fatalism and test anxiety (a measure of the degree to which apprehension about performance on tests and examinations exists) would be high. Moreover it might be expected that performance on these measures and others would vary from one school to another. For a variety of reasons students in certain schools might be expected to manifest less fatalism than others, be less anxious about performance on tests and have higher self-esteem. This was indeed the case. Self-esteem can be taken as an example. The mean self-esteem score for the 1,504 respondents completing the noncognitive tests was 17.60. But there were wide differences among schools in average performance ranging from a high of 21.20 for a highly selective semi-private school (admitting students on the basis of high aptitude) in Addis Ababa to a low of 12.59 for students in one school in the provinces.

Certain measures of social outcomes identified above—for example, fatalism (should be low), self-esteem (should be high), and test anxiety (low)—not only represent social outcomes of education which are ends in themselves but probably serve as proxy measures for a host of personal, social, and motivational variables that affect a student's performance, outlook, and attitudes. That is to say, the social outcome indicators enumerated above, while important in their own right, probably are related to a wide range of attitudes and behaviors. To the extent that this is the case there can be confidence in the fact that the outcome variables have been identified. On the basis of studies conducted in Ethiopia evidence exists that this was indeed the case. Fatalism and test anxiety can be taken as examples. (See Table 9 which presents the results of intercorrelations between the cognitive and the noncognitive measures.) For the 1,504 students referred to above fatalism as well as test anxiety were correlated negatively with performance on a test of aptitude. This means that scores on these measures were inversely related to performances on the aptitude test—the more fatalistic and test-anxious students achieved lower scores.

Table 9. Intercorrelations Among Cognitive and Non-Cognitive Measures; Results of an Ethiopia Study.

	1	2	3	4	5	6	7	8	9	10	11
1. Aptitude	1.00	.16	.22	-.23	-.25	.04	.22	.39	.21	.14	.30
2. General Achievement Motivation	.16	1.00	.46	.07	.19	.29	.26	.14	.16	.11	.13
3. School Achievement Motivation	.22	.46	1.00	.07	.17	.34	.26	.17	.12	.08	.15
4. Fatalism	-.23	.07	.07	1.00	.33	.28	.00	-.02	-.01	.01	-.06
5. Test anxiety	-.25	.19	.17	.33	1.00	.45	-.04	-.16	.11	.00	-.23
6. Anomie	.04	.29	.34	.28	.45	1.00	.16	.02	.07	.10	-.02
7. Locus of control	.22	.26	.26	.00	-.04	.16	1.00	.32	.15	.07	.27
8. Self-esteem	.39	.14	.17	-.02	-.16	.02	.32	1.00	.37	.23	.51
9. Status concern	.21	.16	.12	-.01	.11	.07	.15	.37	1.00	.25	.20
10. Rigidity	.14	.11	.08	.01	.00	.10	.07	.23	.25	1.00	.22
11. Social desirability	.30	.13	.15	-.06	-.23	-.02	.27	.51	.20	.22	1.00

Self-concept and a measure of sense of control of the environment, on the other hand, were correlated positively with the aptitude measure. As with all such measures of relationship it is difficult to know which came first. Is the individual fatalistic and anxious because of low ability or does low ability lead to a fatalistic and anxious individual? Nevertheless the intercorrelations suggest that the social variables studied indeed are not isolated from other educational outcomes, and hence these variables along with others undoubtedly play a key role in a variety of performances and behaviors. Many of the correlations reported in Table 9 appear quite interesting and one might be tempted to engage in ad hoc speculation about their meaning. But the point in discussing the few relationships mentioned is to illustrate that, within the Ethiopian context, the noncognitive measures studied appear to be operating in a predictable fashion. Obviously their future use and value will be greatest when they are tied to the testing of specific hypotheses.

Measuring Entrepreneurial Activity in Developing Countries

After much research, McClelland and others developed the hypothesis that need achievement (*n* Ach) is an element in economic activity of individuals, nations, and communities (McClelland, **45**, 1961). He and his colleagues developed a measure for need for achievement—a way of scoring stories written about some of the pictures used in the Thematic Apperception Test (TAT). Then a count of the number of ideas related to achievement was used as the score of the need for achievement. Several studies were made of achievement levels among groups.

Stories or fantasies were used to demonstrate to what channels people's thoughts turned under pressure or without pressure. A high *n* Ach score was a good predictor of a better performance in a working situation, for one who scored high was more active, more concerned about doing a task better, learned as he went.

Results of an achievement motivation training experiment for adults in a developing country were also measured (McClelland and Winter, **46**, 1969). A cross-section of businessmen important in two small cities in India were trained. In one town, Kakinada, training was given to 52 men in four courses in 1964. In the other, Vellore, 26 men were trained in 1965. Participants' performances before and after training were compared with control groups' performances to determine whether training affected individual effort and community business expansion.

Measurement and evaluation of course results were conducted by interviews and case studies of trainees at six-month intervals over a period of two years after the course. In addition to measuring the outcomes of the training, the interviews and follow-up contacts served as a retraining and reinforcement of former training. Specific questions were asked of participants and the control group members over a period of two years to obtain fairly detailed information about their firm and their role in changes, problems, and plans for the firm.

Trained participants were found to be more active, attempted and actually started more new ventures, mobilized increased amounts of labor and capital, and tended to be responsible for relatively larger percentages of increases in gross income of their firms than were control group members.

The following table shows the increase in entrepreneurial activity (defined as planning and/or taking steps to invest in or establish a business) of participants as compared to control groups (McClelland, **46**, p. 213). Generally those trained were working harder and, in spite of governmental inertia, had more success expanding their businesses than those in control groups.

Percent of Participants who Increased Activity	Before Course 1962-1964	After Course 1964-196
Participants in *n* Ach training	18%	51%
All controls	22%	25%

Increased achievement motivation showed up in the following economic measures:

Hours worked—at the end of the two-year follow-up twice as many participants reported working longer hours after the course than before.

New firms—between 1964 and 1966 almost one trained man in four started a new business (increase from four percent to 22 percent); controls remained at about the same level (seven percent to eight percent).

Capital invested—about one-third of all businessmen made specific investments; after training that proportion rose to three-fourths.

Labor employed—participants employed greater numbers after training than controls.

Support for the "McClelland hypothesis" is by no means unanimous. Other psychologists have noted that the belief that achievement motivation is a key factor in economic development is an oversimplification and that we need yet to understand more fully the cultural, political, economic and individual interrelationships involved in changing society or in modernization.

INDICATORS OF BROAD PATTERNS OF SOCIAL AND POLITICAL CHANGE AND APPROACHES FOR MEASUREMENT

Ultimately attitudinal and behavioral changes of individuals are reflected by changes in the total pattern of social relations. Such changes are envisioned in development aims and point to consideration of the social and political outcomes of education in the broader perspective or terms of global indicators.

Examples of such indicators are those identified by the Ethiopian survey which sought to compile indicators of social and political outcomes of education. Among the indicators identified were:

1) social mobility;
2) social integration;
3) national integration and national unity.

In the Korean study cited above (33) Hong has also identified and made use of pertinent broad phenomena as can be seen from the following examples.

1) technological development and industrialization
2) raising the living standard of the people
3) democratization of political system
4) growth and expansion of middle class
5) rationalization and scientification of life patterns
6) political stability
7) emergence of strong leadership
8) national unification
9) modernization of agriculture

10) long range economic planning
11) full utilization of manpower
12) establishment of social security system

The effects of education on these and related outcomes should be the focus of measurement. Indicators of the type identified in the Ethiopian survey seem to be of particular importance to many developing countries. It may be appropriate to concentrate initially on those outcomes on which education appears to have a more direct influence.

Let us first consider the possible effects of education on *social mobility* in developing countries. One common feature of traditional societies is that status and power are most often inherited, not achieved. The centers of power are generally found among the elite groups, usually small in number, who inherit their social positions. The ranks of the elite groups are not open to the large majority of the population. In other words the opportunity for social mobility is very limited in such societies.

Modern education, "dominated by achievement and universalist norms" (Coleman, 11, p. 17) operates as a factor of *social mobility*. Acceptance of education as an instrument of social change implies a social acceptance of educational achievement as a basis for reward and for ascribing social status. Consequently the expansion of educational opportunity in developing countries could open the door for social advancement for increasing numbers of the population.

Social integration may refer to the degree of "homogenization" that occurs in the pattern of behavior and measured similar (or dissimilar) behavior of members belonging to various cultural groups existing in traditional societies. Dissimilar ethnic, tribal or religious groups may continue to exist side by side in traditional societies with little intercourse or visible influence on each other for many generations, or the sought after integration may be achieved with its objective measurements of communication between groups, visiting between groups, intermarriages, joint business ventures, and so forth.

Related to the process of social integration is the function of education in influencing national integration or national unity. The term "nation-building" is commonly applied in reference to the function of forging unity out of the diversity of relatively isolated elements within a country. The need and desire for national unity is a real one in many developing countries.

In countries where education was applied as a deliberate factor for national integration the results have been remarkable. The cases of Russia,

> The experiences of these countries demonstrate convincingly that educational systems can be powerful instruments in forging national unity, in developing a common language of political communication, and in providing exposure to, if not inculcating a positive effect for, national symbols and goals (Coleman, 11, p. 227).

Japan and the Philippines have been cited as examples in this connection.

In this connection, a potentially strong integrating force in many developing countries is the development and use of a *lingua franca* throughout most of the school system. While this is recognized and is often pursued as a matter of national policy, it may sometimes result in a conflict among language groups involving resistance to one language gaining dominance over others. Nevertheless, the use of a common language or newspaper reading in that language are among the criteria by which national unity may be assessed. Others are intertribal marriages and "visits."

Approaches to Measurement

As already indicated, the identification of social and political outcome indicators is needed to round out the measurements of educational results. For the measurement effort to be useful the indicators must reflect as fully as possible the specific social and political objectives. Several approaches may be taken in measurement. One approach is to focus on changes in individual attitudes. A series of instruments by which to carry out such measurements was mentioned briefly in Chapter 3. Another method calls for measurements based on information on actual behavior, which may prove to be more reliable than that which seeks to tap the predisposition or attitudes of individuals. The rationale for this is that the presumption that changes in individuals are reflected in total patterns of change may stand on firmer ground when based on reported or observed behavior rather than on attitudes. This indeed is the reason that the use of unobtrusive observational measures is indicated.

The nature (not necessarily the substance) of some of the questions included in a comprehensive rural survey questionnaire conducted in Ethiopia by the Institute of Development Research of Haile Sellassie I University may serve as an example (**24**, 1973). Questions such as the following were asked in that survey:

How many times have you used the following facilities in the last 12 months?

	Once	2-5 times	5-10 times	Above 10 times
Bank				
Post Office				
Clinic				
Hospital				
Government schools				
Police department				
Other (specify)				

How many holidays do you observe in a month?

Five to ten days ....

Ten to fifteen ....

Fifteen to twenty days ....

Twenty to twenty-five days ....

Over twenty-five days ....
Don't know ....

Have you ever done anything to try to influence a decision made by local government, and if yes, were you unsuccessful?

Yes ....
No ....
Can't remember ....
Was successful ....
Not successful ....
Can't remember ....

Did you register and vote for the Chamber of Deputies candidates from this district last June?

Register .... yes .... no
Vote .... yes .... no

If reply is no, give reasons ....
....
....

Have you ever contacted your representative in parliament?

Yes ....
No ....

The answers to these questions when correlated with extensive questions on educational backgrounds and socio-economic status serve to gauge educational outcomes.

Unobtrusive measures draw on records pertaining to individuals and may be used to evaluate achievements and competencies accruing from education. Similarly public records which may by design be made to include educational profiles of individuals could serve the several purposes of outcome measurement.

Official statistical data pertaining to particular communities or the nation at large, including trends in demographic changes, health practices, consumption patterns, and so forth, can be related to increases in literacy and the general educational level of the citizenry. This approach may be of particular value in the evaluation of the impact of specific programs (such as the UNESCO-sponsored Work Oriented Adult Literacy Programs) committed to specific objectives and restricted to given communities. Valid conclusions as to the impact of education in the nation at large are more difficult to arrive at on the basis of such studies alone.

Various approaches and methods of measurement (see Table 10) may be attempted in a given situation. This table summarizes the outcome indicators possible in this broad area and suggests methods or approaches of measurement. This is indicated because of the variations of the scope and purposes of educational programs, the difficulty of capturing the intangible outcomes of education, and the multiple factors at work in the dynamic processes of development and change. Here indeed is a challenge for planners and educators, not only to experiment in the use of the methods identified but also to devise other ways by which the social and political impact of education can be better assessed.

Table 10. Possible Approaches for Measurement of Selected Outcome Indicators.

Outcome indicators	**Approaches for Measurement**
"Personal Development" Indicators	
1) Achievement	• Cognitive tests • self-reports • Unobtrusive (non-reactive) measures
2) Attitude towards work, professionalism	• self-report (interview, questionnaire) • employer reports
3) Initiative	• observation (structured) • self-report (interview) • employer reports
4) Leadership	• self report (questionnaire) • unobtrusive (non-reactive) measures (election to office, promotion) • employer reports • non-cognitive tests
5) Motivation	• self-report (interview, questionnaire)
6) Rationality	• self-reports (non-cognitive tests, interview)
7) Risk-taking, "secularization," self-reliance	• self-reports (non-cognitive tests, interviews, questionnaires)
8) Tolerance	• Self-report (non-cognitive tests) • observation • reports of associates
"Social" Outcome Indicators	
1) Social mobility	• self-reports (interviews, questionnaires) • unobtrusive measures (background, income)
2) Social integration	• self-reports, (questionnaire on inter-tribal marriage, associates, language spoken at home) • unobtrusive measures (migration, urbanization data)
3) Social participation	• self-reports (questionnaire on membership in organizations, civic and community participation) • unobtrusive measures (membership, contributions)

4) Health level improvement	• self-report (questionnaire on health practices, attitudes) • unobtrusive measures (infant mortality, incidence of certain diseases)
"Political" Outcome Indicators	
1) National integration	• self-reports (questionnaire on attitudes, political participation) • unobtrusive measures (migration, urbanization)
2) Democratization	• self-reports (questionnaire on attitudes towards justice, merit and reward, voting practice) • unobtrusive measures (voting pattern)

8

8

Nonformal Education and Functional Competence

Nonformal education has been advanced by the International Council on Educational Development for the following reasons: 1) the program can reach into remote areas and further rural development; 2) the programs tend to be inexpensive); 3) low capital and manpower investments involved in nonformal schemes allow them to be altered with a minimum of cost; 4) many new approaches can be developed by educators without the barriers imposed by tradition-bound formal structures; and 5) the functional learning methods to which nonformal programs lend themselves produce skills that directly meet the needs of productive activities. Thus low cost, flexibility, and ease of adaptation for building specific knowledge and skills in a predominantly rural population make nonformal programs more suited to objectives like improving health levels and increasing food production or other productivity. Nearly all nonformal programs are component parts of other development projects, whether sanitation, crop expansion, generation of electricity, in-factory production, or even reinforcing a body politic.

Since nonformal programs can be extremely varied better appreciation may be gained by considering some examples. Some nonformal programs are designed to provide farm families with ancillary skills for home improvement, better farming methods, skills for earning additional income through sideline activities. These are usually rural extension programs. To the extent that they are successful they will contribute to increased productivity and improved standard of living—both of which are development objectives. An example of this type of program is that conducted by the Office of Rural Development (ORD) in the Republic of Korea.

Other programs provide rural young persons with skills for off-farm employment. This type of program is productivity increasing and contributes to relieving population pressures on the farm. Hence it may well contribute to the productivity and well-being of those left behind, as those trained move to other employment. Examples include Kenya's Village Polytechnics, already mentioned, and Thailand's Mobile Trade Training Schools.

Programs are also designed to upgrade and broaden skills of practicing artisans, craftsmen, and small entrepreneurs. These are rural programs aimed at improving the skills of incipient craftsmen and merchant classes. Examples are the Rural Artisan Training Centres in Senegal and Vocational Improvement Centres in Nigeria.

Nonformal indigenous training systems or apprenticeship programs aim at developing self-help skills or new trades. Examples include Accion Cultural Popular (ACPO) in Colombia featuring self-help projects for home improvement, ORD (Korea) for home and community improvement centers, and the previously mentioned Village Polytechnics and Vocational Improvement Centres. These programs can contribute to many dimensions of education and help to reach that large segment of the population which is most neglected in educational development.

An example of a project designed primarily for agriculture, but with a strong education component, is the CADU (Chilalo Agricultural Development Unit) in Ethiopia. Its goals are a) to bring about economic and social development in the project area; b) to give the local population a sense of responsibility for development work; c) to verify methods of agricultural development; and d) to train staff to carry on the work. In the course of accomplishing these goals the project would produce changes in attitude such as would be captured in the Adelman and Morris indicators (1, 1967). The project developed a sense of cooperative effort which contributes to a transformation of the traditional attitudes and functioning of the agricultural sector. The training and attitude changing effects of this project are educational outcomes subject to measurement which contribute to development objectives in Ethiopia.

Most developing countries have extensive health training facilities in recognition of the fact that improvement of health skills and services is an important, if often implicitly stated, goal of development. Health education contributes to an:

1) increase in life expectancy;
2) increase in the quality of life;
3) increase in productivity;
4) decrease in absenteeism;
5) decrease in accidents;
6) increase in family planning; and
7) change in income distribution.

An example of an educational project contributing directly to development on a micro scale level in several underdeveloped countries, Ethiopia in particular, is the Work Oriented Adult Literacy Project (WOALP). This program, which is discussed in detail later, is an integrated development education program incorporating the acquisition of reading, writing, and arithmetic skills, occupational/technical knowledge (productivity increasing courses for agriculture or industry), socioeconomic knowledge, as well as instruction for health, hygiene, nutrition, and changes in attitudes. The basic idea is to teach a gamut of skills in the context of the individual's work and life style.

SOME CONSIDERATIONS OF PLANNING NONFORMAL EDUCATION PROGRAMS

The notion of "nonformal" education should not imply that it can be left to a laissez-faire mode of proliferation. The process of planning for nonformal education differs little from the processes of analysis of more formal structures. Systematic planning and the ability to consider the long as well as short view are necessary and perhaps even more attainable in nonformal processes than formal ones because objectives are more specific and structural change more important to success. Before decision on whether or not to implement a particular nonformal project is made, studies of the following types should be considered: 1) a simple supply and demand analysis, confined to the relevant market (suitable for projects producing specific marketable skills); 2) the number of people served directly (participants) and indirectly (those who receive the impact of what the participants learn); and 3) the availability of manpower to conduct the programs.

Evaluation is, of course, based on project objectives and criteria of assessment that can capture and assess the effects of implementing project objectives as fully as possible. If the project objectives are stated in general terms such as "rural development," for example, these would have to be defined operationally in such terms as a rise in crop yields or food supplies (by some percentage), better nutrition for the population of the nation in terms of daily nutritional requirements, less migration into urban areas, and so forth. It may furthermore be desirable to specify the skills and knowledge that are judged to be important to the achievement of the final project purposes. What must be learned, by how many, and in what setting, in order to carry out the project purposes, would have to be specified.

Nonformal education is usually integrated with other programs; hence costs of providing learning often link outcomes of nonformal and formal methods of education and other development programs as well. There are many respects in which formal educational programs or other projects may be considered as entailing "joint" costs or resources without significant marginal change. Among the resources that may be shared are:

school plant;
school equipment, especially TV;
agricultural agents;
health workers;
school teachers as instructors; and
factory equipment as teaching materials, etc.

This aspect needs to be carefully explored before a decision is made on the implementation of nonformal education. Aside from revealing the cost implications such a study can throw light on the process of development itself in specific situations.

MEASURING OUTCOMES OF NONFORMAL EDUCATION

Outcome measurement is perhaps even more important to nonformal education than to the formal educational system. In the formal educational system there is minimally the yardstick of "years of attendance" and also school leaving examination grades. For the secondary and tertiary levels there is frequently the homogeneous student characteristic that guides evaluation. These are lacking in nonformal education.

In Chapter 9 the question of work skill measurement is discussed. The skill measurement problem is no different in the context of nonformal education than it is in a more formal structure. In both cases it is necessary to know whether the behavior of the individuals who have participated in a learning exercise gives evidence of the learning. Skills acquired, of course, are not the same for all tasks. The worker who operates complex machinery—for example, in an electrical energy generation installation—may require more elaborate measuring rods than a peasant who works his own land with a hoe and has a defined crop yield.

Broadly speaking nonformal education has two types of outcomes: 1) personal development and 2) project development. To analyze the effectiveness of a nonformal program it may be necessary to move back from the ultimate purposes such as increased crop production, or reduced debility (disability), and lower death rates to criteria of measures of specific knowledge about fertilizer, seed, planting, harvesting, or skill in use of farm tools, and so forth. At the same time the need for human development suggests that not only should immediate knowledge and skills be imparted in nonformal programs but that access be kept open to encourage further personal advancement. The midwife who learns about the relationship between hygiene and maternal or infant deaths may prove so competent that acquiring additional health knowledge and skills would be desirable, both in the national interest and in the interest of achieving the initial project purposes (that is, reduced maternal and infant deaths). Assessment of competence and of experience may provide an avenue to more formal education and further personal development.

Personal development objectives may be little different from those of the formal educational system. Of direct project concern is personal betterment through learning, for example, about crop yields, energy generation, family planning, or nutrition. Further progress of the individual marks the follow-on steps. Thus a follow-on of knowledge and skill acquisition that points to a more open system permitting cross movement between the nonformal learning and the more formal methods is necessary. Potential advance is in part necessary to motivate performance in the first instance, but it is also a way to insure that the necessary skills are effectively developed in the appropriate numbers throughout the hierarchy of skills.

To permit such flexibility it is essential that measurement criteria be adopted that permit evaluation by the more formal education system criteria pertaining to knowledge and skills acquired nonformally. The conversion of experience into tested competence will allow crossovers from one system to another.

When nonformal education aims at transmitting specific usable skills, such as skill at farming a crop or automotive repair, a method for measuring the program's success is to use observation tests during and at completion of courses and follow-up with surveys on the job or in the village where the skills are applied. This approach of evaluation certainly has much in common with more formal structures, as do the processes of formulation of objectives and design of criteria of measurement.

In addition to evaluation based on cognitive knowledge and behavior other approaches may involve comparative performance measures and socioeconomic phenomena. Examples of each are given below:

Enrollment in civil service training courses.

Unemployment rates by educational levels comparing unemployment of those with formal schooling and those with nonformal training. This could be done by regions in each country.

Attrition rate of primary and secondary schools compared to attrition rate in nonformal programs.

Percent of persons between the ages of 14-50 enrolled in nonformal education programs.

Illiteracy reported by persons 14 and older.

Annual value of agricultural products marketed in rural sector.

Enrollment in women's organizations according to following categories providing education for:

  Basic health and hygiene
  Dietary and household organization
  Domestic farming.

Morbidity and mortality rates.

Skills, knowledge, and attitudes toward health by "testing" according to age categories. Included may be such items as:

  houseflies can spread serious diseases
  vaccination is the reason so few people get smallpox
  pasteurization of milk kills bacteria harmful to man
  urinating in drinking water may be harmful
  urinating in the moonlight does not cause syphillis
  combinations of foods in choosing a well-balanced meal.

Malnutrition rates over a period.

Crop yields and types of crops.

Infant death rate.

Measurement efforts in nonformal education should avoid the pitfalls of inadequate or inaccurate evaluation in which the wrong questions are raised about outputs and the relation of inputs to outcome. In a case not too atypical, a farmers' training center was analyzed in detail after a reasonable period of time had elapsed to determine whether the courses offered were having the desired impact upon the production behavior of the participants. Due caution was taken to ensure that results were measured against a randomly selected control group. It was found that the impact of the courses was low and that the student farmers were not literate enough to absorb the material effectively. It was recommended that future classes select farmers with a higher degree of literacy, where possible, and that literacy courses be provided before the students begin work at the centers. This case illustrates that, because of preconceived notions about the farmers and the training course, the analysis did not seriously consider the possibility that the materials used by the center for instructing farmers, or even its entire approach to learning, was perhaps unsuitable.

AN EXAMPLE OF EVALUATION OF A PROGRAM IN ETHIOPIA

In this section we will examine the Work Oriented Adult Literacy Project (WOALP), designed to provide functional instruction. We selected this program because it is one of a very few such programs and, more importantly, it has been extensively evaluated.

Its name is somewhat of a misnomer since—while it does provide literacy instruction to adults—it is concerned with far more than academic skills and literacy. As far as literacy is concerned the program centers on reading, writing, and numerical skills in the framework of occupational and living skill development. Some flavor of this project is given by the following descriptive quote:

> This new approach is thus functional in its overall purpose and work-oriented in its specific direction which aims at helping the individual adult in improving his status in the sphere of earning a living as well as his style of life with particular reference to his needs for better health, better civic and social relationships and a richer sense of well-being (UNESCO, **68**, 1972).

Currently WOALP operates in several developing countries, including Ethiopia, where it is located in an experimental agricultural cooperative venture called Chilalo Agricultural Development Unit (CADU). It was set up there since it was felt that CADU would provide a well-defined rural population with well-defined needs as well as an available pool of techniques, implements, etc. However, one of the problems of evaluation is that of separating the education content due to CADU from those due to WOALP.

Within WOALP the participants receive instruction on different types of wheat and barley seeds, fertilizer applications, farm tools adapted to the area, insecticide and pesticide use, and animal husbandry practices through their literacy classes, by demonstration, and also through educational media such as film and radio. Thus practical work skill instruction provides the materials for literacy instruction. Additional material is provided by information on nutrition, hygiene, political processes, and social and cultural practices.

The intended developmental consequences, including economic returns of WOALP, would flow from the expected elevation of the skill level of participants. The measurement of the increase in skills (or lack of it) would constitute a direct measure of the investment outcome of education. Although there is not a scale of output increases due directly to the program, its evaluative questionnaires and interviews provide some indication of its outcomes.

The multifaceted nature of the program gives rise to a variety of indicators of progress and outcome. The evaluation processes of WOALP involve forming indicators and measures of the following categories:

1) Reading: participants are tested on their ability to read passages based on material relating to work and living skills.

2) Writing: once again, material is drawn from everyday life and the test is based on accuracy in writing from dictation. The native language is used in all tests.

3) Basic arithmetic: tests cover the four basic arithmetic operations: addition, subtraction, multiplication and division.

4) Applied arithmetic: use of basic arithmetic in real life situations.

5) Occupational/technical knowledge: questions relate to knowing specific techniques for agriculture and home economics.

6) Socioeconomic knowledge: knowledge of political processes as well as social conditions—including aspects of national history, culture and civics.

7) Health, hygiene and nutrition: questions generally cover practical knowledge taught in women's functional literacy classes.

8) Attitudinal change: scales from test results can cover such specific objectives as acceptance of change, goal-oriented values, work efficiency and community-oriented activities, as well as such general attitudes as activism, participation, empiricism and modernity.

9) Changes in production: respondents are asked to try and evaluate or estimate qualitative and quantitative changes in agricultural yield. The best measures would be based on written records of observation of samples.

10) Changes in number of durable goods used: simple questions are used which ask if there has been investment in better tools or ownership of labor-saving and consumer goods.

11) Changes in income of individuals: this is a simple measure of the change in net wealth due to WOALP.

12) Educative effects on or "scholarization" of participants' children.

The categories of indicators and outcome measures listed above give some notion of the scope of education for functionality in WOALP. In what follows, by way of illustration, more detailed information is given on occupational/technical knowledge and production changes (numbers 5 and 9 above). The evaluation was performed by the administrators of WOALP in Ethiopia, who analyzed responses to three groups of questions by participants. The first group indicates knowledge of improved procedures or tools. (The productivity implications of WOALP can be measured partially by the degree to which participants absorb such knowledge.) The second group relate to whether participants actually used the techniques more extensively than nonparticipants. (This is even a more direct measure of educational outcomes.) The third group require the participants to assess subjectively the changes in personal development and quality of life concomitant upon completion of WOALP.

To disentangle the consequences of different inputs samples were taken of individuals of similar socioeconomic backgrounds in four groups: participants in WOALP who received inputs from CADU (PI), participants in WOALP who did not receive inputs from CADU (PNI), nonparticipants who received inputs from CADU (NPI), and nonparticipants who did not receive inputs from CADU (NPNI). The presence of inputs from CADU means individuals received demonstrations from CADU about seeds, fertilizers, and perhaps tools on credit. Hence the presence of inputs implies a source of education different from WOALP. These four classifications allow the evaluation of the separate impact of WOALP.

The results relating to the first group of questions as described above are contained in Table 11. The four different groups are identified by the letters PI, PNI, NPI, and NPNI. The statistical tests provide an indication as to whether the WOALP had an impact. The response differences which are being statistically tested are characterized by the letters of the two groups, that is, the differences between participants with no input from CADU and nonparticipants with no input from CADU is symbolized by PNI-NPNI. The statistical significance of the difference in responses between pairs of groups is determined by the X^2 test. Accompanying the X^2 values calculated is an indication as to whether the results are statistically significant, and at what level of confidence. For example the results in section A (knowing the improved varieties of wheat recommended for the area) show that among individuals receiving no inputs from CADU, WOALP had a favorable impact. The results in section A also indicate that CADU by itself (NPI-NPNI) had a significant favorable impact. Overall, the WOALP seems to be an effective source of instruction on basic agricultural methods. A possible exception is section F (knowing about recommendations for animal husbandry practices) where the difference in knowledge between participants with inputs and nonparticipants with inputs is not significant. In the WOALP report it was suggested that the impact of CADU was responsible for the lack of difference.

The relation between these results and productivity is that knowing about recommended practices may lead to their implementation and hence demonstrated productivity-increasing effects. The next group results in Table 12 sheds light on the connection between participation and use of recommended practices.

Table 11. Relation Between WOALP and Acquisition of Knowledge*

A . Knowing the improved varities of wheat recommended for the area	
PI - NPI	PNI - NPNI
X^2 = 2.40 (not significant)	X^2 = 7.92 (significant at 1 percent)
PI - PNI	NPI - NPNI
X^2 = 0.66 (not significant)	X^2 = 9.99 (significant at 1 percent)
B. Knowing the improved varities of barley	
PI - NPI	PNI - NPNI
X^2 = 16.26 (significant at 1 percent)	X^2 = 13.16 (significant at 1 percent)
C. Knowing about fertilizer application	
	PNI - NPNI
	X^2 = 6.72 (significant at 1 percent)
D. Knowing any recommended farm tools introduced by CADU	
PI - NPI	PNI - NPNI
X^2 = 20.30 (significant at 1 percent)	X^2 = 20.62 (significant at 1 percent)
E. Knowing about recommendations for insecticides and pesticides	
PI - NPI	PNI - NPNI
X^2 = 6.50 (significant at 1 percent)	X^2 = 7.92 (significant at 2 percent)
F. Knowing about recommendations for animal husbandry practices	
PI - NPI	PNI - NPNI
X^2 = 0.92 (not significant)	X^2 = 7.56 (significant at 1 percent)

* (Singh, 61, 1972, pp. 8-17.) For a description of these symbols see the text.

The first two sections, A and B, indicate the efficacy of WOALP concerning its investment outcomes. In both sections the differences between participants and nonparticipants are significant, though at differing levels. Section C (using fertilizer as stated by the respondents) however, does not indicate any significant differences. Two explanations are that the practices engendered by CADU spread to both participants and nonparticipants equally and hence both groups realize the value of and make use of fertilizer, and that WOALP just was not successful in this instance.

The results of Tables 11 and 12 indicate that the participants in WOALP acquire and use knowledge which has proven productivity consequences.

Table 12. Relation Between Knowledge and Implementation of Improved Methods of WOALP in Ethiopia*

A. Using some improved tools as stated by respondents

PI - NPI

$X^2 = 11.04$ (significant at 1 percent)

B. Using some insecticides or pesticides as stated by the respondents

PI - NPI

$X^2 = 5.16$ (significant at 5 percent)

C. Using fertilizer as stated by the respondents

PI - NPI

$X^2 = .08$ (not significant at 5 percent)

---

* (Singh, 61, 1972, pp. 8-17.) For a description of these symbols see the text.

Hence such questionnaire techniques can be used to measure the investment outcomes of educational projects such as WOALP. Similarly, the subjective assessment by the participants and the control groups provides a direct measure of the outcomes of this project.

One problem with analysis of a project like WOALP is that it is not clear which facet of its educational product is reflected by the measures, since WOALP is a multipurpose program. It could be the literacy aspect, the joining of literacy and work orientation, or there may be a demonstration or Hawthorne effect. Of course the program's concept is that all these parts work together more effectively than they would separately. Yet it is unclear if the results could not be produced from just one of the parts.

9

Measuring Work Skills

Work skill training encompasses programs that teach, in some structured manner, capacity necessary for skilled, semi-skilled service, clerical, and agricultural occupations. The term "structured" implies programs constructed around a curriculum, involving personnel trained or selected for teaching and subjects, and hence programs, which lend themselves to evaluation.

There are a great many programs in the developing world which offer work skill training. A survey of nonformal education programs can be found in International Council on Educational Development (**37**, 1973). The distinction between work skill programs and all nonformal education programs is best captured by noting that work skill programs are composed of vocational training while nonformal education may include nonvocational training such as health and hygiene instruction or types of literacy instruction.

One example of work skill programs is Kenya's Village Polytechnics which seek to provide rural youth with employable skills in urban areas. Another example is the Vocational Improvement Centres in Nigeria, instituted to help broaden or upgrade the skills of craftsmen, artisans, and entrepreneurs.

Most of the examples in this chapter involve testing materials from the developed countries—particularly the United States—simply because the most detailed and thoroughly validated type of measures are available in these countries. Exceptions include the work of the International Labour Organisation in developing and evaluating vocational training programs in many developing countries and the work of the Ministry of National Community Development and Social Affairs of the Ethiopian Government cited below.

Just as in adapting aptitude and achievement tests from developed countries, adaptation of work skill tests must proceed cautiously. For discussion of problems of adaptation of aptitude and achievement tests see Chapters 2 and 3. Obvious problems in intercultural translation must be solved. But work skill testing may be more amenable to this type of adaptation than the traditional achievement testing. Occupational testing relies heavily on paper and pencil tests and hence on the language skill problem. Furthermore occupational testing in all its forms is narrower in scope and concerned with practical and well-known jobs and competencies. The use of work sample tests when well constructed involves testing with less cultural bias than other forms of examinations. Work sample testing involves knowledge acquisition and also manual skills whose possession is noncultural.

Nevertheless care must be taken with work skill tests, as with other types of tests adopted from other sources, to make them not only as culturally fair as possible but structured at a level consistent with the requirements of the economy and the nation and region. When effort is made to adapt tests to the specific culture and job skill level requirement of the country, the results are improved. For example in tests used to predict work efficiency in a West African mine it was found by Dawson (17, 1963) that test results were most highly correlated with work performance for tests developed in terms of the indigenous culture.

Unlike achievement and aptitude testing, education and acculturation have little effect on manual skills (see Ord, 53, 1972).

FUNDAMENTALS OF WORK SKILL MEASUREMENT

Discussed here are some of the important concepts and considerations which go into the design, administration, scoring and use of performance tests to measure the outcomes of vocational training. As a first step designing vocational education requires investigation to determine the level of skill competence in certain trades to be taught. This information is usually obtained from consultation with experts or by job analysis. Of the two, job analysis usually provides more accurate information.

Obtaining information from experts is less costly than through job analysis, but may be inadequate or inaccurate. These shortcomings may result from the peculiarities of duties in a given location; recent changes in duties of which the expert may not be aware; and the tendency of human beings to emphasize those aspects of a job they consider most important or typical. Use of expert opinion alone is most useful when on-the-spot job analysis is not practical because of job location, time limitations, or a lack of skilled analysts.

Job analysis, a systematic listing and examination of the duties involved, is done through observing people at work and interviewing workers and supervisors. The nature of such information should be related to the purpose for which it will be used. Job analysis for test construction purposes should report the relative frequency and importance of a task, its difficulty, the tasks which require the greatest skill, criteria of proper performance, special tools and equipment needed.

Once test criteria are obtained the next step is to prepare for its administration. All good tests, no matter how administered, have one common characteristic—objective measurement. To achieve this certain precautions must be taken. For example, questions must be stated the same way to all examinees and testing conditions such as room lighting, noise level, writing facilities and available tools and equipment must be as uniform as possible.

Ideally, a test should provide a complete picture of the particular aptitudes, knowledge or skills being measured. Since such a complete evaluation is impractical, tests rely on representative sampling. Certain principles in selecting test content help achieve representativeness. Most important is that the selected material be essential, not merely desirable. An automobile mechanic could properly be tested on his ability to identify and repair a defective carburetor rather than on his knowledge of the relative advantages of four versus two-barrel carburetors.

It would usually be more appropriate to test an automobile mechanic on his ability to tune an engine, align front wheels, or replace brake linings than on his ability to rebore engine cylinders, which might be the proper domain of an engine specialist.

A good test should involve more than one task, for example tuning an engine. There is no specific rule for the number of tasks which should be measured to provide adequate representation. Expert judgment and, if practicable, experimental validation of the test can best provide such information.

Tasks chosen for a work skill test should include at least a few that separate the competent workman from the man meeting minimum standards. A novice automobile mechanic would be expected to tune an engine in a routine "tune up." It requires more skill to tune an engine if its distributor has been improperly replaced.

It is important also to consider the practicability of including particular tasks in the test. Among the pertinent considerations are the length of time necessary to complete the task, the requirement for specialized tools and equipment, the ease of objective scoring and the possibility of injury to an unskilled or nervous examinee. For example, overhaul of an automatic transmission in an automobile would not ordinarily be a practicable task in a test situation because of the time and expense involved. Demonstration of the manner of checking transmission fluid pressure would be preferable.

Communication is a vital consideration in testing. No test can measure competence unless the examinee clearly understands what is expected of him. Instructions must be clear, in the right language, and should use the same terminology as in the locality in which the examinee is trained.

TYPES OF WORK-SKILL TESTS

There are three basic types of work-skill tests: oral, written, and work sample. Each type of test is discussed below:

Oral Tests

Oral tests are commonly used for individual testing. They are used a) when written tests would be too costly to prepare, b) when the level of reading ability of the examinee is low or questionable, or c) when the manner of oral response by the examinee is important. Employee selection through interviews by one or more "examiners," if carried out objectively, is a type of oral testing.

Much can be learned about "exposure" to and understanding of a trade through oral tests. As developed, such questions are not designed to grade the relative competence of knowledgeable craftsmen. They may, however, be properly used to demarcate proficiency.

Oral interview tests can be used if the examinees have difficulty in reading, when there are too few examinees to warrant preparation of a costly written or work-sample test, or when the ability of the examinee to respond orally is a factor to be rated.

Since oral tests are usually used for selection purposes involving relatively few individuals, per-capita cost considerations normally preclude the kind of statistical objectivity exercised in the preparation of written or work-sample examination. Therefore special care should be taken in the selection of questions to see that the material covered is important and representative. Since the inherent rigidity of the written word is absent, examiners should be particularly careful to ensure that the procedures followed and the language used are the same for all examinees. Use of a commonly understood language is particularly important if the examinees come from varied cultural and linguistic backgrounds.

To limit the inherent tendencies toward subjectivity in oral examinations, procedures and questions should be written out and a specific scale designed to evaluate responses.

Certain kinds of subjects such as the use of tools and the meaning of technical terms lend themselves readily to oral examination. Such questions are easy to prepare and can be objective and valid.

Technical experts can give assistance in the preparation of oral trade questions and can be used as examiners. If the time and facilities permit it is highly desirable to try out oral interview questions on skilled personnel of known competence prior to their actual use on examinees.

The U.S. Training and Employment Service of the Department of Labor has tested individuals in over 700 different occupations. Although discontinued after World War II, they could provide a solid basis for the development of modified versions in many less developed countries.

The oral trade questions used by the Employment Service were designed to elicit short specific answers. To limit guessing, any questions which might be answered by "yes" or "no" were omitted. Every effort for clarity and relevance was made in questioning, and the questions were developed by studying job content, questioning qualified workers, and discussing with supervisory employers.

An initial group of questions for each trade was tried out on three different groups of workers: a) experts, b) apprentices and helpers, and c) workers in related occupations. Those questions which differentiated individuals in group a) from those in b) and c) were retained; the others were discarded. The retained questions were edited and revised; the final test of their effectiveness was made on the comparable three groups throughout the United States. At least 50 experts of the first group and 25 each of the others were tested in the final nationwide tryouts.

Through statistical analysis of the trial data, a final set of questions (usually about 15) was selected. For each test a score range of three levels was provided—well informed, some information, and little information.

To illustrate the questions, a set used for bricklayers is included below.

Table 13. Oral Trade Questions—Bricklayer (Examples)

What do you mean by "building up a lead"? (pronounced "leed")
Building up end section (corner) of wall.

What kind of joints do you use when plaster is to be put directly on the brick?
Cut off (rough) (vee).

What kind of course do you use in brick structure to bind the front and back walls together?
Header course.

What do you call a brick that is laid on the narrow edge?
Rowlock

Interpretation	Score
Well informed	60-100%
Some information	20- 59%
Little information	0- 19%

Written Tests

The most widely used test is the written test because they may readily be prepared, administered and scored objectively under standardized conditions. Large numbers of individuals can easily be tested at one time—and inexpensively. The test administrator or proctor needs to know very little about testing and even less about the technical content of the test. Good written tests are available for a wide variety of purposes, including measuring general mental ability, specialized aptitudes, knowledge of academic subject matter and job related information, spatial relations ability, and art judgment. No specialized equipment is needed for written tests.

Their major shortcoming is that they do not permit an assessment of on-the-job skills on a relative basis. A craftsman may know all about a job and still produce an inferior product or work at a slower than acceptable pace.

Written tests may be used in some instances as complete or partial substitutes for work-sample tests. They can be used as complete substitutes if the duties of the job concerned involve written output as they do for professional and semi-professional occupations. They may also properly be used to determine knowledge which is job-related or can ordinarily be acquired only by on-the-job experience. But for most skilled trades written tests are ordinarily poor substitutes for actual performance tests.

Written tests can determine the extent to which an individual has been exposed to a job. A carpenter should know the normal height of stair risers or the vertical parts of a paneled door. But such trade information might not separate a carpenter from a lumberyard salesman or knowledgeable homeowner. Questions must involve knowledge of special tools and equipment, working procedures and technical terms to be useful for a particular skilled occupation.

Written tests can also be used to assess knowledge required for job performance. For example a carpenter could be asked to calculate the amount of studding lumber required for a room. He could also be asked how he would bore a hole at a 45° angle through a 2x4. Likewise, an electronics technician could be asked to specify how he would determine the wave pattern at a certain point in a television circuit.

Probably the most extensive store of written test questions exists in the U.S. Department of Defense. The military services classify personnel into about 1,500 different occupational specialties. Although many of these specialties are peculiar to the military system, most have counterparts in the civilian economy.

Military job types have been aggregated into nine major occupational areas of which seven have civilian counterparts. The written tests prepared for jobs in five of these areas have considerable applicability for testing similar civilian jobs. The job structure listed in Table 14 differs from the civilian counterpart principally in the absence of tasks concerned with production and the high degree of specialization in the electronics maintenance and repair occupations.

Work-Sample Tests

Work-sample tests are usually the preferred measuring tool for selecting skilled craftsmen. They inherently minimise the need for language facility and in this respect are particularly suitable for use in developing nations. Furthermore they are superior to oral or written tests for work skills in that they measure competence to do certain tasks rather than to exhibit knowledge considered related to the tasks. But work-sample tests are most costly to prepare and more difficult to administer, particularly for large groups. They often require special shops, tools and equipment, and a large number of examiners.

The overriding advantage is that they possess more "face-validity" than do written or oral tests. Face-validity is a term used by testing experts to describe test questions which are obviously related to their purpose.

Table 14. Military Occupational Areas with Civilian Counterparts.

Electronic Equipment Repairmen

Includes the maintenance and repair of radio, radar and electronic navigational, countermeasures, missile guidance, torpedoes, nuclear weapons control, equipment control, equipment and related or allied equipment maintenance.

Medical and Dental Specialists

Includes patient care and treatment, technical and related medical and dental services.

Other Technical and Allied Specialists

Includes technical and professional-type skills not classified elsewhere such as survey, drafting, photography, cartography, meteorology, scientific aides, physical laboratory analysis, ordnance disposal and music.

Administrative Specialists and Clerks

Includes personnel and administration, supply and transportation administration, accounting, communications center operation, and related activities.

Electrical/Mechanical Equipment Repairmen

Includes maintenance and repair of electrical, mechanical, hydraulic and pneumatic equipment.

Craftsmen

Includes forming, fabricating and installing structures and components; utilities installation and maintenance; operation of construction equipment and power tools and related crafts and trades.

Service and Supply Handlers

Includes protective and personal services, non-clerical personnel involved in warehousing, food handling and motor transportation.

Validity is achieved by selecting test content important to a job which, in fact, discriminates between the competent and incompetent worker. To ascertain whether a test meets these criteria it is necessary not only to use the judgment of experts but also to give experimental tests.

The procedures used in the experimental administration of a test must be as close as possible to those to be used in the actual routine use of the real test. The kinds of examinees must be similar and the tools, time, equipment, shop facilities, and scoring methods the same.

Valuable experimental techniques include follow-up on the examinees to see how well the test predicts on-the-job performance or, in certain experimental designs, how well the highly skilled man on the job compares with the less skilled individual or the novice in the test.

It must be emphasized that a work-sample test is only as good as the care exercised in its preparation to meet the criteria of validity and reliability. Such a test constructed without strict adherence to the principles of good test construction could be less valid than a properly designed written or oral examination. A valid performance test is one which accurately predicts on-the-job competence regardless of who rates the examinees, and it must give consistent results if administered to the same examinees at different times.

As an example of the manner in which a work sample test may be prepared, the procedures used in developing a test for Machinist's Mates (MM) in the U.S. Navy are summarized below.* Only one of the work samples in the test is used to illustrate the procedure.

A Machinist's Mate in the U.S. Navy operates, maintains and repairs steam propulsion and auxiliary equipment. He is not a machinist.

After a review of job analysis material and consultation with senior Navy MMs, the test designer decided that the most frequently performed and critical tasks that a journeyman MM had to do involved the ability to:

1) identify and use hand tools common to the job;
2) perform maintenance and read meters and gauges accurately;
3) make simple repairs in pressure lines;
4) perform simple troubleshooting and systems analyses in pressure systems;
5) operate certain equipment; and
6) assemble and disassemble "high frequence" failure systems.

To sample duty No. 2 above, a test was designed to check the individual's ability to cut a gasket used in piping systems.

The instructions given the examinees are to cut a gasket. The tools furnished are: a flange, a ball-peen hammer, asbestos gasket, material and some bolts. The time allotted for completion of the task is 10 minutes. The scoring checklist used by the examiner shown below is scored on an all or none basis for each rated factor.

Other material included in the test measures the examinee's knowledge of the names of tools used making and breaking a flange, his ability to read certain meters and to diagnose the fault in a pumping system.

As noted previously, objective scoring of multiple-choice written tests is easily achieved. In some types of work-sample lists objective scoring is also simple to accomplish. In others, subjective judgments must be made, but the human factor of examiner judgment should be limited as much as possible.

In the gasket-cutting test most items can be scored objectively. There is some room for difference in judgment as to when a hammer is used "lightly" or whether there are "scratches in the final product."

In tests involving construction of objects or repairing machinery it is practicable to measure size, closeness to tolerances or precision of angles

---

* The material on the MM test is based on a report of the Office of Naval Research, dated June 1972, titled, "Nonverbal and Culture Fair Performance Prediction Procedures," authors: Arthur I. Siegel and Brian A. Bergman.

Table 15. Scoring Checklist for Making a Full-Face Gasket.

1. Uses round end
2. of ball peen hammer . . . .
3. and correct size gasket material . . . .
4. and taps out one bold hole . . . .
5. lightly . . . .
6. Inserts bolt in hole . . . .
7. Taps out diagonally opposite bolt hold and inserts bolt . . . .
8. taps out remaining bolt holes in any order . . . .
9. lightly . . . .
10. Taps out inside circumference of flange . . . .
11. using round end of ball-peen hammer . . . .
12. lightly . . . .
13. Taps at outside circumference of flange . . . .
14. using flat end of ball-peen hammer . . . .
15. lightly . . . .
16. Remove tapped out pieces of gasket material from gasket and flange . . . .
17. No frayed edges in final product . . . .
18. No gouges or scratches in final product . . . .
19. Does not ruin or throw away any pieces of gasket material . . . .

at joints. Scores can be set at the various levels of precision in the final product. In tests involving location of defects or "trouble shooting" of systems, ratings can be given in terms of time required to identify the source of the difficulty. The nature of the work sample will test the ingenuity of the test designer to develop appropriate objective testing criteria.

Undoubtedly the low profile of vocational education, and the cost of developing performance tests, have been principal factors in hindering the growth of vocational testing beyond the written examination stage. Unfortunately in most cases written tests do not make good substitutes for practical ones. A good craftsman is not necessarily competent with the written or even the spoken word. He is best measured by what he can do, not by what he can write or say.

The difficulties in constructing and administering performance tests discussed earlier have been a major hindrance to the production of such tests. Few work sample tests have been adequately standardized and validated. While at this writing the number of available good work sample tests is low, it is quite likely that the near future will bring considerable improvement.

In mid-1973, under the sponsorship of a consortium of 44 states, a National Occupational Competency Testing Institute was established at Princeton, New Jersey. Operating under the long-established, prestigious Educational Testing Service, the Institute plans to make major inroads into the area of performance as well as written testing for occupational competence.

The Institute was set up as a consequence of the research project mentioned above. The project reviewed the national need for occupational

Table 16. Tests Available from the National Occupational Competency Testing Institute

Aviation Industry Occupations
Airframe and Power Plant Mechanic
Automotive Industry Occupations
Auto Mechanic
Auto Body Repair
Diesel Engine Repair
Small Engine Repair
Building Industry Occupations
Air Conditioning and Refrigeration
Carpentry
Electrical Installation
Masonry
Plumbing
Sheet Metal
Drafting Industry Occupations
Architectural Drafting
Machine Drafting
Electrical Industry Occupations
Industrial Electrician
Electronic Industry Occupations
Communications Electronics
Industrial Electronics
Food Industry Occupations
Quality Food Preparation
Graphic Industry Occupations
Printing
Machine Industry Occupations
Machine Trades
Metal Industry Occupations
Welding
Personal Service Industry Occupations
Cosmetology
Technology Industry Occupations
Civil Technology
Mechanical Technology
Wood Industry Occupations
Cabinet Making and Millwork

testing, investigated the "state of the art" in this area, and experimentally evaluated the feasibility of establishing a consortium to prepare occupational competency tests.

While the primary focus of the Institute's effort is the preparation of competency tests for vocational/industrial teachers, the tests and associated materials developed by the Institute are useful for assessing the competence of workers in the skilled trades in general.

As of 1973, the Institute had available tests in the industries and occupations listed in Table 16. These tests consist of both written and performance material. Although not all of the categories are appropriate for less developed countries, many of them could be adapted.

Work sample tests have also been used in some of the less developed countries to measure competency. An example is provided by the National Industrial Vocational Training Center established by the Ministry of National Community Development and Social Affairs of the Imperial Ethiopian Government.

In that program participants are required to perform normal occupational skills under careful test conditions. For example, the carpenter trainee is required to demonstrate his ability to construct joints—as well as other tasks. Figure 1 is an example of a work sample test commonly

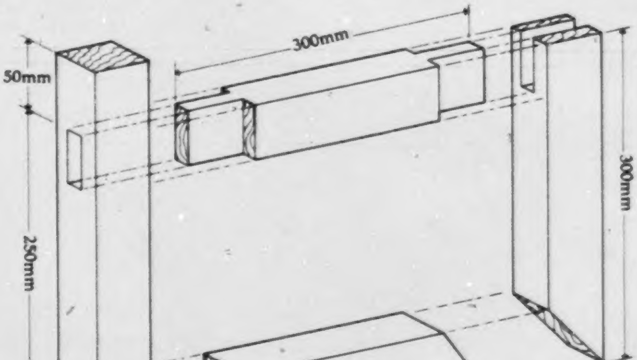
Figure 1 CARPENTRY

(Joinery - test)

Make following joints according to the sketch.

Use 4x5 planned wood, glue, screws and nails.

Follow the measurement and the connection will be exactly square-formed.



employed in this program. The individual is given the tools and materials and is required to construct joints of the specified dimensions within a specified time limit. Successful completion of the tasks indicate real job competency.

There are work skill tests in use there for carpentry, automechanics, plumbing, electrical equipment repair, and other skills.

10

The Economic Outcomes of Education

Economists have done extensive work on defining and measuring the effects or outcomes of education. The results developed are macroeconomic and microeconomic in nature. The methods and problems of these measures are reviewed in this chapter.

MACROECONOMIC MEASURES OF EDUCATIONAL OUTCOMES

In the past, emphasis has been on the measurement of education's contribution to total economic growth. Foremost among these contributions have been the works of Denison and Schultz in the United States. Schultz addressed himself to capital formation through education. Denison essentially reduced the overall growth rate into portions attributable to physical capital formation, growth in the labor supply, technological improvement, and the improvement in the productivity of labor due to education. These types of analyses do not measure the educational outcomes of particular types of nonformal education.

Another approach to the measurement of economic outcomes of education calculates the marginal or average productivity component of specific forms of education—such as formal education or various types of nonformal education. The productivity component or measure is the contribution to economic output of an educated person due solely to the individual's training. The measurement of educational productivity may be direct—the actual change in output due just to education—or indirect—that is, through the use of an intermediate measure of productivity such as income changes due to education. The indirect measures are the most common; many studies for a variety of countries feature the calculation of the rate of return or the net present discounted value of education—the two most widely used indirect measures of the investment outcomes of education.

Education can produce changes in productivity or economic growth through changing technology by fostering invention and innovation. This dynamic effect of education can be measured indirectly, though imperfectly. We discussed this measure in Chapter 9. In the next section, a brief description of Denison's approach will be given, and the basic results examined.

"RESIDUAL" APPROACH TO EDUCATIONAL OUTCOME MEASUREMENT

Denison assumes that education affects aggregate growth through its effects on labor quality (18, 1964). In his framework, education can increase labor productivity because 1) an educated person is likely to perform better than an uneducated person, 2) education leads to greater individual flexibility and initiative, 3) education implies a better labor market information system and thus more efficient labor allocation, and 4) growth increases need for technologically trained individuals familiar with basic techniques and amenable to modern working conditions.

Denison's analysis uses the concept of an aggregate production function to separate the relative contributions of labor, land and capital to GNP. The particular form used in these studies is the Cobb-Douglas production function which can be specified:

$$1)\ Q_t = e^{\phi t} A_t^{\alpha} L_t^{\beta} K_t^{\gamma},\ \alpha + \beta + \gamma = 1$$

where Qt is output in period t, At^{α}, Lt^{β}, Kt^{γ} are land, labor, and capital inputs in period t, e is the natural constant, and ϕ, α, β, γ are parameters.

1) Measuring the productivity of a given increase in education by relating increases in education (years of schooling) to income increases. This implies, as will be explained in a later section, equating income changes with productivity changes. It also implies correcting for extraneous factors which may lead to increases in income but which are not a result of education (for example, ability and family background). This is done by arbitrarily assuming 60 percent of income difference correlated with education are a result of productivity increases.

2) Adjusting results to allow for the effects of longer school years over time, since this has been the trend in both developed and less-developed countries.

3) Finally, using information on the change in the educational attainment of the labor force and the effect such changes produce in productivity to compute the increase in labor productivity over time. An increase in labor productivity is equivalent to an increase in the labor force itself in this method and can, therefore, be used to calculate education's aggregate contribution to increases in GNP.

Summary results of Denison's study are presented in Table 17. They show the effects of education on two measures of aggregate U.S. economic performance over time: the growth in total real national income and the growth in real national income per person employed. The results demonstrate the importance of education's contribution within the context of the limited model employed by Denison. Education's contribution to the total growth in real national income ranges from 12 to 23 percent and to the total growth in real national income from 29 to 42 percent depending upon the period chosen.

Table 17 Estimates of the Contribution of Education to Past and Future Growth of Real National Income, U.S.A.

	1909-29	1929-56	1960-80
Growth of total real national income	2.82%	2.93%	3.33%
Amount of growth rate ascribed to education	0.35	0.67	0.64
Percent of growth rate ascribed to education	12	23	19
Growth rate of real national income per person employed	1.22	1.60	1.62
Amount of growth rate ascribed to education	0.35	0.67	0.64
Percent of growth rate ascribed to education	29	42	40

Source: Denison (**18**, 1964, p. 35).

There are three major criticisms of these macroeconomic measures of educational outcomes. First, the use of income differences to reflect productivity differences is inaccurate due to the inapplicability of the marginal productivity theory of income distribution and the effect of other factors besides education on income. Second, many economists believe that the aggregate production function does not accurately reflect the production process of an economy which includes so many heterogeneous inputs, or that the Cobb-Douglas form is not the correct specification of an aggregate production function. Finally, there is much debate as to whether the "residual" approach can explain aggregate growth, or whether there is any reason to favor the educational "explanation" of the residual over another.

DIRECT MEASUREMENT OF THE PRODUCTIVITY EFFECT OF EDUCATION

The productivity increasing effect of education can be measured by relating changes in physical output to changes in educational status. The change in educational status can be due to formal education or specific, skill-increasing nonformal educational programs. The measurement of the productivity increasing effects of formal education are more difficult to measure directly since much of the effects of such education cannot be directly related to work skills or future employment. Indirect measurement as covered in the next section is a useful way around this problem.

For nonformal education programs the direct measures consist of identifying the work output and measuring the change in that work output as only one factor changes—the education of the work force. In manufacturing processes an inservice educational program may lead to improved skills, organization, and morale and the work yield may increase. In practice there are difficulties in holding all other influences constant.

One example of direct productivity measurement is provided by the Work-Oriented Adult Literacy Project (WOALP) in Ethiopia—which was discussed in Chapter 7. As part of the evaluation of this project participants and nonparticipants were asked to evaluate the changes in yield over the period of time of instruction. Table 18 is drawn from the evaluation study. As described in Chapter 7 the education project took place within the Chilalo Agricultural Development Unit (CADU) and hence there was the problem of disentangling the effects of WOALP and CADU. This is the reason for the classifications PI (WOALP participant with input from CADU), NPI (non-WOALP participant with input), PNI (participant in WOALP with no input), and NPNI (non-WOALP participant without input). Once again the presence of inputs means use of seed types, common agricultural equipment, perhaps even demonstrations.

Table 18 The Overall Assessment of the Use of Improved Agricultural Practices—A Subjective Assessment by the Respondents

	PI		NPI		PNI		NPNI	
Responser	n= 72		n= 72		n= 36		n= 36	
Yield up 75-100%	48	66.67%	40	55.56%	11	30.55%	7	19.44%
Yield up 50-75%	13	18.06	12	16.67	13	8.33	2	5.56
Yield up 25-50%	—	—	7	9.72	2	5.56	3	8.33
Less than 25%	—	—	9	12.50	3	8.33	3	8.33
Not increased	3	4.17	—	—	7	19.44	8	22.22
Decreased	4	5.76	2	7.78	1	2.78	4	11.11
Not stated	4	5.76	2	2.78	9	25.00	9	25.00
TOTAL	72	100%	72	100%	36	100%	36	100%

Source: Singh, **61**, 1972, pp. 8-17.

The results of the questionnaire or interview procedure can be used in their present form to measure educational outcomes. They can also be converted into alternative, more useful, forms. An example might be using the results of Table 18 to construct expected increases in yield for each of the form groups. This implies summarizing each of the columns into single numbers expressing the expected yield for each of the four groups.

If additional data were collected more direct and objective assessment of the consequences of education could be made. For example, if observations were made of the use of the techniques taught by WOALP, and these were correlated with observations of the actual increases in yield due to the different procedures, then we would have a direct and objective assessment of the educational outcome of WOALP.

Table 18 shows mixed results. The comparison favors WOALP when there are no inputs from CADU; in the case of inputs the results are less clear. The point however is that, depending on the accuracy of subjective evaluation, sample survey methods such as this are a way of measuring the direct productivity consequences of nonformal education programs.

MICRO INVESTMENT MEASURES

One of the best known methods for analyzing investment decisions is cost-benefit analysis. By comparing the costs and benefits of investments in any form—whether it be plant, equipment, inventories, or education, a variety of measurements of the value of an investment can be formed. In this section those measures, as applied to human capital formation, and examples of their use will be explained.

Cost Benefit Analysis

Economic markets typically perform functions similar to those of benefit-cost type analyses—they direct the allocation of funds into purchases of goods and services. The types of decisions that benefit-cost analysis may help make, however, are not generally amenable to market solutions; education is one area where the market system does not perform its allocation function efficiently.

The study of capital formation through education is very similar to the well-developed theory of physical capital formation. Investment in education and training is, like investment in physical capital, a process which yields benefits over time in the future and which may also produce costs in the future. It is a different process than the buying and selling of goods in the present.

The heart of benefit-cost analysis lies in a clear and accurate description and measurement of the benefits and costs associated with education.

The Cost Elements

The costs of education are somewhat easier to conceptualize. The direct costs of education are those costs incurred in the provision of education, including:

1) salaries of teachers and auxiliary staff;
2) supplies provided by and for the school;
3) rent for land usage;
4) interest (debt service).

The indirect costs are those incurred by students (or the school system if the student is subsidized) in pursuing his education. These include:

5) room and board if the student has to live away from home;
6) additional clothing required for school;
7) transportation to and from school;
8) cost of personal supplies.

Finally, there are costs which are measured by the foregone opportunities of pursuing education instead of its alternatives. Examples of opportunity cost include:

9) the value of the tax exemption enjoyed by non-profit educational institutions; and
10) the opportunity cost of the student's time (employment forgone) while in school.

These categories include the most commonly mentioned cost items; however, in actual measures of costs many items are excluded or only roughly approximated for lack of data.

A distinction can also be made between social and private costs of education. Social costs are those borne by society in providing educational facilities; they are covered by all of the categories listed above. Private costs are those items that the individual personally pays. These include many of the items listed, such as transportation or personal supplies, as well as that portion of the other costs he may have to pay under the heading of tuition or fees. Since the social cost category is more extensive, social costs are equal to or greater than private costs. This distinction becomes important later in analyzing the private *versus* social investment component of education.

In many studies of the costs of education the largest category is the opportunity cost of the student's time. This can vary widely from one educational level to another; it is lowest in the primary level and highest at the university level. At least one study has found that opportunity costs represent 50 percent of total resource cost for four years of university education. In practice usually only the direct costs and opportunity costs are captured. The former is sometimes easily available in the form of total educational expenditures by a ministry of education. Opportunity costs are approximated by the salaries earned by persons with a lower educational background. For example the opportunity cost for a high school student is the average wage earned by an individual with a primary school education. In some cases other costs may be imputed—for example, personal supplies may be imputed as 10 percent of individual costs.

One example of cost data is provided by the study of Kenyan education by Hans Thias and Martin Carnoy (64, 1972), which made use of widely scattered figures to calculate direct costs. Total teacher's salaries were computed by multiplying the average salary for each category (primary school teacher with two years post primary school training, for example) by the total number of teachers in that category. To estimate expenditure on equipment the capitation grant was multiplied by the total number of students. Other expenses were approximated by a lump-sum expenditure by the Ministry of Education. Capital costs were estimated by a set sum per pupil based on a certain number of places provided by a school building of a specified type. For example, the capital costs for a primary school with a permanent structure was estimated at K£ 12-20 per pupil, while the capital costs of a traditional "mud-and-wattle" type structure were only K£ 1 per pupil. Because school buildings have different lifespans, the costs were amortized over 40 years for the permanent structure and over 10 years for the traditional structure.

For illustration it might be useful to give another example, drawn from a benefit-cost analysis of the Ethiopian educational system, which illustrates the problems that occur and means of overcoming them. Table 19 summarizes total costs per student by education level. Direct costs were defined to include recurrent costs (teacher's salaries, maintenance, student out-of-pocket costs, etc.) as well as capital costs. Wastage captures the repetition and failure costs inherent in the educational process. (Earlier we commented on this wastage concept.) Income foregone is the opportunity cost of the students' time and is approximated by the average wages earned at the lower educational level.

Table 19. Total Costs Per Student by Educational Level in Ethiopia

	Direct Costs	Wastage	Income Foregone	Total Cost Social	Total Cost Private
Primary (6 yrs.)	$ 468	$ 132	$ 309	$ 909	$ 417
Junior secondary (2 yrs.)	288	—	675	963	693
Senior secondary (4 yrs.)	1,060	324	2,677	4,061	2,787
Tertiary non-degree (2 yrs.)	5,400	810	3,181	9,391	3,075
University (4 yrs.)	17,408	6,093	6,826	30,327	6,641

Source: Planning Commission Office of Ethiopia (8, 1973).

The direct costs are gleaned from three sources. The first is national educational expenditures by the state on a given level of education divided by the number of students at that level. The Ministry of Education and Fine Arts has been the main source for this type of data. Capital costs were estimated at eight percent of total costs at the primary level, 18 percent at the junior secondary level, 21 percent at the senior secondary level, and 20 percent at the tertiary non-degree and the university levels. Out-of-pocket costs for students were assumed to amount to 10 percent of total costs at the primary level and $50 annually for university students.

Estimating problems are many—both conceptual and factual. One problem with this study was that a large proportion of students attend non-government schools having much lower costs per pupil. Yet the cost per student figures were taken from government sources. The solution was to estimate national costs per pupil as a weighted average by enrollments in the different types of schools, using available data for the level of cost discrepancy in the various types of schools.

Table 19 shows that there is considerable discrepancy between social and private costs, particularly at the higher levels of education. Social costs are higher because of government subsidization of direct costs and the inclusion of wastage in social costs. Private costs were the additional direct costs borne by the student (very small at all levels), and income foregone minus any subsistence allowance.

The Benefit Elements

Educational benefits are of many kinds but only a few are quantified for inclusion in the analysis. A partial listing of the private and social benefits as usually enumerated for this purpose is given below. The private benefits include:

1) increased earnings for the individual;
2) consumption benefits for the individual;
3) increased capacity of individuals to adjust to new circumstances, jobs, job opportunities.

The social benefits include:

4) provision of manpower for economic growth;
5) increased productivity for society;
6) benefits to society of invention and innovation that are generated by increased levels of education;
7) better citizenship (and hence lower crime costs);
8) intergenerational effect;
9) external effects on other individuals, family, employees, and employer; and
10) discovery and cultivation of talent by the educational system.

The standard procedure for estimating the productivity consequences of education is to assume that the marginal productivity theory of income distribution holds, and then to use the direct link it provides between income and productivity. That theory states that under conditions of perfect competition an individual is paid the value of his marginal product of education. We use the change in income level an individual received as a result of the change in his educational status as a proxy for the increase in productivity the change in education may occasion.

In practice the procedure is to use income levels across educational classes as a measure of the investment benefits of education. This requires detailed census or questionnaire data which are usually difficult to obtain for developing countries. The problem is compounded by the need for information on more than just education and income data. Many other factors affect income levels besides education—a person's ability, his family connections, his age or years of experience, his occupation or where and for whom he works. Data must be collected for all of these variables in order to isolate education's contribution.

Nevertheless such specialized data are collected and have been used in a number of developing countries. The Kenyan study used results of cross-sectional urban earnings data collected in the Labor Force Survey made in January-February 1968 and a rural cross-sectional survey of households in the Central Province in 1963-64 (Thias and Carnoy, 64, 1972). The Labor Force Survey consisted of 4,742 interviews collected in Nairobi, Mombasa, and Nakuru. Questions were asked about wages or salaries, sex, years of schooling completed, type of school, age, father's occupation, parent's literacy, ethnic origin, number of years spent in the city where interviewed, age of finishing school and beginning work, number of years with present employer, whether or not currently receiving formal education, whether or not the individual received on-the-job training, if African, his or her tribe, whether or not the individual belongs to a union. The less extensive rural survey generated a sample of 800.

The next step is to generate age-income profiles by means of regression analysis, using an equation of the following form:

$$(2)\ Y = B_0 + Bi\,X + \sum_{i} Bi\,Zi + U$$

Where Y is the income variable, X is the age variable, and the Z's are socioeconomic and occupational variables. This type of equation is specified for each educational class. The results capture the effect of education without the influence of the other factors affecting some income classes.

Table 20. Age—Earnings Profiles for Kenyan Males, by Years of School, 1968, Adjusted for All Variables*

Years of Schooling				Earnings in Kenyan Shillings per month			
Age	0-2	3-5	7F[a]	7P[b]	7a11	9	11a11
2-14	—	—	289	—	287	—	—
15-16	—	—	183	—	266	—	—
17-19	261	264	267	320	243	310	125
20-24	332	275	276	322	266	185	385
25-29	339	328	339	410	345	315	484
30-34	385	353	388	558	497	384	791
35-44	387	387	399	596	466	685	898
45-54	380	383	505	907	633	539	405
55+	379	358	—	592	460	1981	—

* These results are corrected for experience, ability, family background, ethnic background, economic status.

[a] Failed Kenyan Primary Examination (KPE).

[b] Passed KPE.

Source: Thias and Carnoy (64, 1972:43).

Turning to the Ethiopian example, a similar procedure was used with an equation of the form:

(3) $\text{Log } Y = B_0 + B_1X_1 + B_2X_2 + B_3D + U$

Y is the annual cash income in Ethiopia; X_1 is age, X_2 is years of schooling, D_1 is the dummy variable used to distinguish the effects of public versus private sector employment. The data were collected by the Ministry of National Community Development and Social Affairs in 1971-72 for a sample size of 476.

This equation is a different functional form from that used in the Kenyan study. Here educational attainment is captured by direct inclusion of the education variable; the numerous socioeconomic occupational variables of the Kenyan study are absent (except for public versus private sector employment) due to lack of data. Also the semi-log form of equation (4) provided a better fit to the Ethiopian data.

The actual estimation for the Ethiopian data produced the following results:

$$\begin{aligned}(4)\ \log Y = {} & \underset{(0.01787)}{2.36039} + \underset{(0.00052)}{0.01135X} + \underset{(0.00096)}{0.08025X} \\ & \underset{(0.00997)}{-\,0.13765} + u\text{ , } R^2 = .945\end{aligned}$$

The numbers in parentheses are the standard errors of the regression coefficients. All are significant at the one percent level. This, in addition to the high R2, has been interpreted as giving great accuracy. The results can also be expressed in the equivalent form:

$$(5)\ Y = \frac{229.30\ (1.0265)^{X_1}(1.2029)^{X_2}\ e^{u}}{(1.3729)\ D}$$

This provides the following information:

1) given the starting income of 229.30, each year of age (experience) increases income, on the average, by 2.65 percent;

2) each year of school raises income by 20.29 percent; and

3) employees in the private sector receive an average of 37.29 percent higher salaries than individuals in the public sector.

Equation (5) provides for the construction of age-income profiles for different educational classes. Yet there are several complicating factors. First, ability, family background, and so forth, are not represented; thus part of the income differences shown between classes is due to these factors, not education. Economists in the past tried to compensate for this problem. Denison, as mentioned, arbitrarily attributed 60 percent of the income difference to education and 40 percent to other factors.

Becker used a sliding adjustment of 60 percent for primary and secondary levels and 66 percent for the tertiary level. In the Ethiopian study the adjustment factors were .5 for primary, .6 for secondary, and .7 for tertiary. These were considered reasonable in light of the international evidence.

The second adjustment needed is for mortality. Future incomes should be adjusted for the possibility of mortality by use of life expectancy tables. In the Ethiopian study this was not done, rather each individual was assumed to work to retirement at age 65. This omission, however, may not make much difference in quantitative terms since mortality is relatively low when individuals are young and their earnings are discounted only slightly, and mortality is high only when individuals are near the end of their working life, and their incomes are discounted heavily.

A third complication is that the investment consequences of education occur only if individuals enter the labor force. Therefore income levels should be adjusted for labor force participation rates. In the Ethiopian study a bias was introduced by the use of participation rates from the National Sample Survey of rural males with the income data for urban males. This is a typical example of problems resulting from lack of data in developing countries.

A fourth adjustment is required for unemployment. Survey data gathered from the employed yield higher income differentials. This correction was not incorporated into the Ethiopian study, which again introduces a bias as unemployment among secondary school leavers appears to be growing into a problem of significance.

Finally, to distinguish between social and private benefits, private benefits are based on post-tax incomes while social benefits are based on pretax incomes. The idea is that tax revenues are used for government expenditures which, presumably, benefit society. If this is the only adjustment undertaken to distinguish private from public benefits, social returns will always be higher.

After taking account of all of these adjustments the benefits of education can be constructed for any age level, as Table 21 demonstrates. There both income levels and income changes between educational classes are captured for age 38. The adjustment for taxes is introduced in later calculations.

The Kenyan study attempts to correct the benefit stream for mortality, experience, socioeconomic variables (previously mentioned), ability, the individual's occupation, unemployment probabilities, and the income effect of the results of the Kenyan Primary Examination—an important examination for further schooling and employment. The exact results will be considered later with the rate of return analysis, but it may be useful to look at the procedures used to adjust for several of these complications. Age-earnings profiles of Kenyan males corrected for socioeconomic, occupational, and ability variables are presented in Table 20.

Table 21. Median Income Estimates (age 38), Ethiopia, 1971-72

	Gross Income	Gross Increment	Net Income	Net Increment
Unschooled	499		247	
Primary	1,510	1,011	746	499
Junior Secondary	2,185	675	1,295	549
Senior Secondary	4,575	2,390	2,712	1,417
Tertiary non-degree	6,620	2,045	4,577	1,865
University degree*	9,579	5,004	6,626	3,914

* Compared with senior secondary

Source: Planning Commission Office (1973:6).

The effect of ability, mentioned previously with respect to the Ethiopian results, can be approximated by comparing the returns of those who pass the KPE (Kenyan Primary Examination) with those who fail. The results of these examinations are not the best measure of ability but some aspects of ability are captured.

Outcome Measures Based on Benefit-Cost Analysis

Benefit-cost analysis can be used to measure the economic outcomes of education. The net present discounted value (NPDV) and the internal rate of return (IROR) are both applications of benefit-cost analysis and indirect measures of the productivity consequences of education; they represent the indirect relationship, under the marginal productivity theory of income distribution, between education and real output.

The NPDV and the IROR are the main measures of the investment outcomes of education. The question is which should be used. The measures are similar and usually produce the same hierarchy of results, that is, investment outcomes in one country will generally, under both measures, be either inferior or superior to those of another country. Yet in some situations the projected investment outcomes differ according to the measure used.

Policymakers have used both measures in considering educational investment. Some calculate the NPDV of all investment projects and use the results to rank the projects and allocate funds accordingly—the higher the the NPDV, the better the investment opportunity. The calculation, of the NPDV, of course, requires a specified interest rate. Alternatively, policy makers may calculate the IROR for all projects and reject those whose IROR is below the going interest rate.

Both measures employ the benefits and costs as described in the last section in their calculation. The benefits and costs cannot be compared directly, for there is a fundamental problem of timing. The costs will be incurred over a short span of time while training benefits will extend well into the future of the individual's working life. Thus the problem is one of comparing costs and benefits over different time periods. Here is how this problem is solved in the economic context. A stream of benefits or costs is discounted in present value terms. Economists recognize that a dollar received in the future is somehow different from a dollar held today. The most obvious difference is that a dollar today can be invested to yield a dollar plus interest in a year. So the problem really is one of comparing values from two different time periods. The solution is to convert future receipts or payments into their present dollar equivalents through discounting. Once the dollar receipts are both expressed in present day value they may be readily compared. Suppose an individual wished to convert $100 received in a year to present value form. If the interest rate is six percent, discounting yields $100/(1+.06) = $94.34, which is the net present discounted value (NPDV) of $100 received in a year. An individual would prefer to have $100 today rather than $100 in a year, but he would prefer $100 in a year to $90 today if he were influenced by strictly economic motives.

Using the NPDV, benefits and costs are discounted over their life at an appropriate rate of interest (how this interest rate is chosen will be discussed later) and are then used to calculate the NPDV of an education investment project according to the formula:

$$(6)\quad NPDV = \sum_{t=1}^{n} \frac{Bt}{(1+i)^t} - \sum_{t=1}^{n} \frac{Ct}{(1+i)^t}$$

where Bt is money benefits in time t, Ct is money costs in time t, and i is the interest rate. The time span is n years—perhaps 45 years for most human capital investments. This formula gives the entire value of the investment in human capital in present dollar terms. Due to the restricted way benefits are incorporated, as discussed in the last section, the money benefits are equivalent to excess earnings due to additional education or training. Also, costs will generally be discounted over a shorter period of time than benefits, as costs for education are generally incurred in four to eight year intervals while benefits may last many years into the future.

As an example of the calculation of NPDV, assume training costs are $1,000 in the initial period but there are no costs after than. Assume further that the investment will yield $100 a year in higher earnings over 20 years (n = 20). Finally, let the prevailing rate of interest be five percent. The resulting NPDV is:

$$NPDV = \sum_{t=1}^{20} \$100/(1.05)^t - 1{,}000 = \$1{,}246.20 - \$1{,}000 = \$246.20$$

This final figure is one measure of the annual investment outcome of education.

The second major investment outcome measure is the internal rate of return (IROR). The formula for the calculation of the IROR is similar to that for the NPDV, but in this case the interest rate is a variable. That rate is determined on the assumption that the benefit stream is equal to value of the cost stream.

$$(7) \quad \sum_{t=1}^{n} \frac{Bt}{(1+r)^t} - \sum_{t=1}^{n} \frac{Ct}{(1+r)^t} = 0$$

To find the IROR, we solve (7) for r, the only unknown in the equation. Referring back to the initial example, we have:

$$\sum_{t=1}^{20} \$100/(1+r)^t = \$1{,}000$$

The solution is r=7.75 percent. This means that over the entire 20-year period of the educational investment, the compound rate of return on the initial investment of $1,000 is 7.75 percent.

EXAMPLES OF THE USES OF NPDV AND IROR

The key issue is to determine what interest rate should be used for discounting in the NPDV for comparing with the IROR. The solution is to use the interest rate that represents the opportunity cost of the investment funds. For the individual decisionmaker that rate might be the yield on long-term government bonds since not having invested his money in his own education, he might choose the alternative use. For society, the rate should reflect the opportunity cost of funds drawn from the private sector and society's rate of time preference between present and future social consumption. In both methods an interest rate must be used either for purposes of discounting as in the NPDV, or to provide a point of comparison as with the IROR.

Thus we must look to other criteria to determine which measure to use. A cogent reason for preferring the NPDV is that it allows comparison between educational investment projects of different time spans, whereas the IROR requires that the time spans of the two projects be the same. Further, the NPDV will reflect differences in the size of educational investments while the IROR will not. The conclusion seems to be that when the IROR and NPDV yield different rankings of investment measures the NPDV is to be preferred.

Although the NPDV is conceptually a better measure of the investment outcomes of education, the IROR is most often used. The reasons are: first, the two measures will generally yield similar rankings of projects; and second, the IROR is unit free (the NPDV is expressed in monetary units, the IROR is in percentage) and thus provides an easier frame of reference for comparisons. In the examples which follow, both measures will be provided where they are available. The first is provided by the

Ethiopian study discussed extensively in the last section. Table 22 shows NPDVs and IRORs for various types of returns to education.

Table 22. Returns on Educational Investment, Ethiopia

Type of Return	Years of Schooling 6	8	12	14	16
Social IROR	20.3%	28.6%	18.7%	14.5%	9.4%
Private IROR	35.0	36.7	22.8	33.4	27.4
Gross private IROR	58.5	47.9	35.4	39.6	34.2
Social NPDV* (mil.)	$120.4	$216.5	$ 97.0	$ 44.6	$ 7.3

* The interest rate used is 10 percent.
Source: Planning Commission Office (1973:10).

The pattern of educational outcomes is fairly typical of many developing countries: high rates of return for the lower schooling levels (primary and junior secondary) and lower rates for the higher levels, the university in particular. The distinction between the social and private internal rates of return derives from the differences between social and private benefits and costs already discussed in the previous section. The category "gross private" is based on the benefits which are used privately but without the correction factors for ability, background, labor force participation, and so forth.

The NPDVs reflect the same pattern of outcomes as the various IRORs. This measure in Ethiopia was based on an interest rate of 10 percent, the relevant return for social investment projects in Ethiopia. The negative returns for university education are the result of a low IROR and the 10 percent interest rate used for the NPDV. The overriding impression from these first results is that investment outcomes of education are extremely high. However caution in accepting these results is suggested by problems discussed later.

The next example, the Kenyan study, is particularly interesting because it attempted to deal with many of the problems associated with benefit-cost analysis. The rates of return as reported in Table 23 give a detailed breakdown of the private rates of return to all levels of formal education. The breakdowns include adjustments for age (experience), taxes, socioeconomic variables, and ability as approximated by exam scores. Furthermore adjustments were made for employment probabilities for the combined urban-rural calculations.

Table 23. **Average Private Internal Rates of Return in Kenya to Schooling, All Adjustments, by Years of Schooling, 1968**

	Urban Rates Corrected for:			corrected for:	
	Age Only	Age and Taxes	Age, Taxes, Soc-econ, var, and Exam Score	Employment probabilities only	Employment probabilities and soc-econ. variable
Primary					
2-4	26%	26%	31%	19%	23%
5-7	55	55	13	17	7
2-7	33	33	18		
Secondary					
8-9	24	24	9		
10-11	52	40	30		
8-11	36	32	19		
Higher secondary					
12-13	24	23	23		
University					
14-17	27	20	20		

Source: Thias and Carnoy (**64**, 1972:91).

As discussed, the adjustments for age, socioeconomic variables, and ability are to separate those non-educational factors which affect the benefit stream. Taxes are adjusted to allow for separation of the social and private rates of return. Employment probabilities are adjusted for the effect on earnings of unemployment in the various educational classes; obviously to consider all individuals as fully employed would bias the rates of return upward. Mortality is not included as an adjustment factor since the individual in assessing his investment return does not consider his lifespan; he can circumvent the quantitative consequences of an early death through insurance.

The pattern of results shows that the rates of return decrease as more and more extraneous factors are taken into account through adjustment factors. The most accurate assessment of the investment outcomes of education is shown under the last column for urban rates where the rate for primary school is 18 percent, for secondary school 19 percent, and for the university 20 percent. There is a general equality of return across educational classes. The private rates are high and demonstrate, as they do for

Ethiopia, that education produces large investment returns for the individual.

The social IRORs to education in Kenya are shown in Table 24. As explained above, the difference between social and private costs is the subsidy element reflected in social but not private costs. The difference between social and private benefits can be measured by the difference between pre- and post-tax income. The returns reported in Table 24 incorporate the distinction between pre- and post-tax incomes and also mortality and government employment effects. Mortality is included since the loss through death affects society's investment outcomes; government employment rates are included since differences between government and private wage rates which may reflect non-productivity differences should not be included in calculations of the rate of return.

Once again rates of return decline as more and more of the extraneous factors are considered. The most accurate assessment of the social investment outcomes of education are under the last column of urban rates where the social return to primary education is 14 percent, to secondary 15 percent, and to university 9 percent.

PROBLEMS IN THE USE OF BENEFIT-COST ANALYSIS

The results reported for both Ethiopia and Kenya demonstrate comparative results between social and private investment outcomes of education which are present in all studies of this type. In both cases the private outcomes are greater than the social outcomes. This is because the social costs included subsidies by the state to education which the individual does not pay; subsidies raise social costs at the same time as they lower private costs. On the benefit side the difference between social and private benefits is interpreted in the narrow sense of the difference between pre- and post-tax incomes. There is some dispute as to whether this difference should be recognized since individuals presumably receive benefits from their tax dollars.

The difference between social and private benefits does not reflect the external benefits of education as discussed in the last section. Thus, it is likely that social rates of return are seriously understated, and would give an inaccurate description of the social investment consequences of education. There has been one attempt to estimate the degree of bias inherent in the usual method of calculating the social return to education. If the social rate of return is interpreted in the narrow sense described, introducing approximations for some of the other social benefits of education, including the external productivity effects of supervisory manpower, can lead to a doubling of the social rate of return. This provides a wide range within which the actual social rate of return may be found.

There have been studies of the rate of return to education performed for many of the developing countries. A complete listing is contained in Table

Table 24. Average Social Internal Rates of Return to Schooling in Kenya, All Adjustments, by Years of Schooling, 1968

	Urban Rates Corrected for				Combined urban—rural corrected for:	
Years of schooling	Age only	Age and mortality	Age mortality and soc-econ. var.	Age, mortality soc-econ. var., ability Kpe Cert. and Govt. empl. effects	Employment probabilities	Employment probabilities and soc.econ.var.
Primary						
2-4	16%	15%	11%	15%	14%	14%
5-7	38	38	17	14	11	5
2-7	22	21	14	14		
Secondary						
8-9	16	15	19	6		
10-11	34	34	28	21		
8-11	24	24	23	15		
Higher Secondary						
12-13	15	15	15	15		
University						
14-17	9	9	9	9		

Source: Thias and Carnoy (64, 1972:92).

25. Only social rates of return are included in the table. In general there is a decline in the rate of return as the level of education increases. Overall, the rates of return to education are high, indicating large investment outcomes of education. They compare favorably with rates of return to physical capital formation.

Table 25. Social Rates of Return to Schooling by Country for Various Years, Primary, Secondary, and University Levels of Schooling

	Year	Primary IROR	Secondary IROR	University IROR
Puerto Rico	1960	21%	22%	16%
Mexico	1963	25	17	23
Venezuela	1957	82[a]	17	23
Colombia	1965	40	24	8
Chile	1959	12	12	9
Brazil	1962	11	17	14
S. Korea	1967	12	9	5
Israel	1958	16	7	7
India	1960	20	13	13
Malaysia[b]	1967	9	12	11
Philippines	1966	8	21	11
Ethiopia	1971	20	19	9
Ghana	1967	18	11	16
Kenya	1968	22	20	9
Uganda[c]	1965	66	50	12
Zambia	1960	12	—	—

a. The 82 percent represents the rate of return between illiteracy and six years of schooling, a rather broad interval. Income foregone is assumed to be zero. These two facts imply that the IROR is seriously overestimated.

b. Rates shown here are underestimates of the unadjusted rates, since they have been corrected for non-schooling factors.

c. Rates for Uganda are seriously overestimated, since they are based on differences between average incomes of employees with different amounts of schooling. This assumes that income differences are constant over lifetime.

Sources: George Pascharopoulos and K. Hinchlitye, "Rates of Return to Investment in Education and the Impact on Growth and Development: An International Comparison," London School of Economics, Higher Education, Research Unit, 1970 (mimeo), and Ethiopian Planning Commission Office (1973).

NPDV and the IROR are subject to two criticisms in addition to those concerning the benefit and cost calculations. First, in analyzing education the choice of units, usually years of education, leads to measurement error due to its impreciseness and discrete jumps (especially at the end of schooling levels such as primary or secondary). The result of such an error is that the rate of return tends to be underestimated. The second problem is that rates of returns based on annual salaries ignore the question of hours worked per year. If education induces an individual to work longer, then the usual rate of return measures will overestimate his investment outcome. The size of these biases or the relation between them is not known with any degree of accuracy.

There is a final problem associated with using the IROR or NPDV as measures of the investment component of education. Since education costs money now and yields benefits in the future, individuals often wish to borrow to pay their current costs when their incomes are low and pay off their borrowings when their incomes are high in later life. To the extent that they are unable to absorb as much education as they desire, the private rate of return to education will be underestimated. This is because an individual's ability and education interact in affecting income. If an individual undertakes less education than he desires or than is efficient, the increment to his income is less by a factor that accounts for the interaction between his education and ability. This shortfall of income represents a real loss in productivity, since the individual with additional education would be more productive because of his additional education and the interaction factor.

This problem results from the imperfect capital markets which often exist in underdeveloped countries. Capital markets require information regarding the returns to educational investment that is notoriously imperfect. Additionally, there is always the risk of default by individuals which introduces an element of uncertainty in the analysis. These factors combine to produce higher borrowing rates (if loans can be had at all), less investment in education, and an understatement of the investment return to education.

The intergenerational investment impact of education provides another measure of educational outcome. The intergenerational investment impact is the increase in earnings over costs by an individual due to the educational level of parents. Many studies have found that a child's degree of education is positively correlated with that of his parents. Indeed parental income and family characteristics are important explanatory variables. This may happen for several reasons. First, better educated parents tend to instill in their children an appreciation of learning; second, children of better educated parents have higher achievement motivation; third, children of better educated parents are more aware of the financial implications of education; and finally, hereditary ability plays a role. Whatever the reason, the result is the same—more education and a greater productivity.

Measurement of the intergenerational impact of education makes use of the methodology of internal rate of return analysis. The social benefits of intergeneration investment are the present value of the extra lifetime earnings for each educational level multiplied by the probability of attaining the educational level actually attained due to parental education and discounted back to the time of the parental education. Estimates have shown that the intergenerational investment outcome of education is large.

INCOME REDISTRIBUTION EFFECTS OF EDUCATION

Education can affect the distribution of income in two separate ways. The first is via the productivity increasing effects of education. This exhibits itself first by increased direct measures of output and then by the additional income that results. The second is through the indirect measure of productivity—the increase in income due to education. Both of these approaches have been documented and examples given earlier in this chapter. The second way education contributes to income redistribution is by the change in the income profile due to financing education.

The key to relating changes in the distribution of income to education in the first instance is to measure the opportunity of acquiring an education by income class of families. The types of direct and indirect measures of productivity demonstrate the rewards associated with education. The degree to which individuals have the opportunity to enjoy these rewards as related to their family income level measure the changes in income profile that can be expected.

An example of relating the opportunity to participate in the income increasing effects of education by income class is contained in Table 26. This table, adapted from a study by Jean-Pierre Jallade, shows the enrollment rates by income class of families in rural Colombia. This information combined with measurements of the income producing effects will, when measured over time, give a rough measure of income redistribution effects of education.

Table 26. Enrollments by Level of Family Income: Rural Colombia

Income bracket (pesos/year)	Enrolled in Primary Education %	Number of children (5-25) thousands
0-6000	15	2305
6000-12000	23	1630
12000-24000	33	843
24000-60000	50	261
60000-120000	100[a]	36
120000-240000	100[a]	5
Over 240000	100[a]	30
TOTAL	23	5100

[a] The 100 percent enrollment ratios for the three upper income groups imply that the children beyond primary school age in those groups have migrated to cities to get secondary or higher education, and are no longer included in their original consumption tax paying unit.

Source: adapted from Jallade (38, p. 32).

The second way education can affect income distribution is by changing the income profile as a result of financing education. This involves measuring the proportion of tax monies returned to various income classes as a result of the public subsidization of education. The greater the proportion of tax monies returned in the form of subsidies the better off, in financial terms, is the family. Table 27 shows, for Colombia as a whole, the ratios just described for primary and secondary schools. These results show that for the primary level the poorest families are the prime beneficiaries of the public school system. The poorest 40 percent receive 87 percent of their taxes back in the form of subsidies. These figures are derived from enrollment figures, taxes received, cost subsidization by the state—all by income classes.

The figures for secondary level show the richest 13 percent and poorest 40 percent contributing to the middle 47 percent. In this case the income changes due to financing education favor the middle classes.

Table 27: Public Subsidies for Three Levels of Education as a Proportion of Taxes Distributed among Income Groups (%): Colombia

Income Bracket (pesos/year)	Number of Households	Public Subsidies for Primary Education as Proportion of Taxes	Public Subsidies for Secondary Education as Proportion of Taxes
0-6000	19.0	109	9
6000-12000	20.2	77	4
12000-24000	24.9	49	18
24000-60000	22.9	22	20
60000-120000	8.8	4	7
120000-240000	3.4	1	3
Over 240000	0.8	—	1
Total	100.0	16	9

Source: Jean-Pierre Jallade, *Public Expenditures on Education and Income Distribution in Colombia*. World Bank Staff Occasional Papers No. 18, 1974, p. 40.

11

Educational Planning and Evaluation

Many nations are finding that their "inherited" educational systems are ill-adapted for the challenges of growth and development. Often existing education tends to be more academic than practical, more elitist than democratic, more urban-oriented than rural; and the systems are criticized for accentuating the inequality between regions, tribes, and religious groups.

This has been the result, in part, of adoption and expansion of educational systems developed elsewhere. Further, when most developing countries began planning for educational development they again came under the influence of the international community, which emphasized expanding educational systems and increasing enrollments geared to meeting skilled manpower requirements. Consequently attempts were made to develop forecasting methods and statistical data for that purpose. Yet the educational system remained basically unchanged. Now however the trend is shifting away from a primarily linear expansion approach. Planners are turning more and more to a search for better use of existing resource commitments through basic changes in the educational system itself: curricular reforms, innovations in teaching methods and improved management. This shift from a quantitative to a qualitative approach reflects changing views of the role of education in the developing countries and is buttressed by the experience showing that traditional European education is not necessarily appropriate education for the developing countries.

The new trends include emphasis on education for rural and agricultural as well as for modern and urban development, incorporating into planning all types of education: formal, nonformal, and traditional—public and private. Hand in hand with this realistic appraisal of educational requirements is a reinforced awareness that education (whether knowledge, skill or attitudes) cannot by itself bring about transformations in a nation—that, for example, rural and agricultural development problems cannot be solved by education alone. Instead there must be simultaneous changes in the environment. The best educational system cannot keep rural residents in rural areas unless the environment is as desirable and attractive as other areas, particularly the city. Educational policymakers and planners, therefore, are becoming increasingly aware of the limitations of the role of education in a nation's overall development.

There is today greater appreciation of the multifunctional role of education in development and an urgent call on the part of educational policymakers in developing countries for improving planning capability which provides for an assessment of how well the educational system is fulfilling its role in influencing development

STATUS OF EDUCATIONAL PLANNING IN DEVELOPING COUNTRIES

Concern for improving educational planning and development is reflected by the efforts of many developing countries. This is highlighted by a brief survey of the experiences in educational planning in recent years in the three continents of Africa, Asia and Latin America.

A review of educational planning by UNESCO in the latter part of the 1960s showed that 80 to 98 countries surveyed had drawn up educational plans (UNESCO, 68, 1970). However, 11 of those plans were not adopted, and three approved plans were not put into operation. Eighty-six of the countries had plans for economic and social development; in 70 cases education was made part of the development plan. In all countries primary and secondary school levels were included in the plans. A number of countries' plans included technical and vocational training; fewer included agricultural training at the secondary level; and many included higher education. While only 59 countries included qualitative aspects of planning (structure and content of education), most others were in the process of developing those qualitative aspects. Sixty-five of the plans were drawn up after an investigation of manpower needs.

Eighty-five percent of the countries had an administrative authority responsible for educational planning. The planning organization was located in various ministries or agencies of government. A number of countries had established permanent planning units. The planning units in some cases dealt only with quantitative aspects of planning while in other cases they had responsibility to coordinate programs of various departments involving all aspects of educational planning.

In the African continent UNESCO and the Economic Commission for Africa sponsored the Addis Ababa conference of 1961 which listed targets for educational expansion. A series of regional conferences in the 1960s were all concerned with various aspects of educational planning: the 1962 Paris conference suggested ways to reduce costs of secondary schooling and made recommendations for better planning and organization; the Tananarive conference of the same year dealt with higher education; and in the previous year the Abidjan conference had recommended that planning cover adult literacy education and scientific research. A general assessment of education in Africa was made by the Nairobi conference of 1968.

In spite of poorly developed administrative structures for carrying out educational plans and severe financial restraints on educational growth, a number of African countries are implementing plans. Among the many problems facing planners are high unemployment among school leavers and differences in educational opportunity both between sexes and between rural and urban areas.

At least one African nation has undertaken a comprehensive review of its educational needs. Ethiopia's Education Sector Review, issued in August 1972, sets forth educational objectives in the context of overall national goals and makes basic proposals for all aspects of education along with projected financial resources. (Report of the Education Sector Review, 23, 1972.) The general objective in the review according to which alternative strategies were calculated was to achieve universal primary education in Ethiopia by the year 2000. Three different strategies were formulated in order to achieve this objective and the participants of the review conference agreed on a combination of two of the strategies.

It is interesting to note that reorienting attitudes was one of the many recommendations made in the Report of the Education Sector Review in Ethiopia.

> Perhaps the greatest challenge in implementing the recommendations of the Sector Review—or any other far-reaching change in the nation's educational system—would be the need to reorient the attitudes of the people of the nation toward the purposes of education. . . . the recommended system would provide a self-contained program at each level that would be terminal for most students. Its adoption would require recognition and acceptance of this by pupils, parents, teachers, educational administrators, and people from all facets of Ethiopian society (p. VII-4).

The long-term objectives of this review have to be further integrated into the country's five-year development plans. It is quite apparent that implementation will encounter a good deal of technical, administrative and financial constraints in view of both the massive expansion and the reorientation of the educational system envisaged.

Again, a series of UNESCO-sponsored regional conferences were held to stimulate the use of planning by Asian nations. The 1960 Karachi conference plan, endorsed by 17 Asian nations, dealt with primary education and set a goal of seven years of free compulsory education to be reached by 1980. The Tokyo conference of 1962 reviewed problems in the Karachi plan and explored development of private and public education in the framework of planning for social and economic development. Three years later the Bangkok Conference adopted a draft model for Asian educational development for the period 1965-1980.

In India and Pakistan extensive and fairly sophisticated planning has been undertaken. India has had a series of development plans since 1951; Pakistan since 1958. In both nations, however, the directions of development have been influenced greatly by social and political pressures and demands. In India, for example, rapid increase in the number of educated persons resulted not so much from planned targets based on estimated projection of demand but from pressures from groups all over India, especially women.

In Latin America the Organization of American States (OAS) and UNESCO have sponsored educational conferences to establish quantitative targets with recommendations for increasing technical, scientific and vocational education. In 1958 an inter-American seminar examined problems of organization and methods of educational planning, administration, and finance; in 1962 the Santiago de Chile conference studied relations between education and economic, social, and demographic conditions in Latin America and defined objectives of a 10-year plan for educational development; and in 1966 the Buenos Aires conference dealt with the question of content, methods, yields and evaluation of education.

Most nations in Latin America have established planning organizations and plans have been developed. There has, however, been difficulty in implementing those plans. Political instability, adherence to the status quo within the educational system itself, and inefficient administrative structures have been cited as basic reasons for ineffectiveness of plan implementation. In many instances, too, planning for education has not been coordinated with general development plans although in Venezuela, El Salvador, and Uruguay, educational plans are concerned with employment and development needs. Latin American countries are faced with such rapid population growth that education has to expand rapidly in order to stay even with the birth rate. Approved educational plans are usually conservative and planners are more inclined to accept the status quo of educational systems than to deal with basic changes in the system.

SCOPE AND PROCESS OF EDUCATIONAL PLANNING

Educational planning is still a developing art, and this is especially true in the developing countries where data limitations are often severe and developmental goals in general, and educational goals in particular, are not clearly defined. Furthermore social and political pressures everywhere make difficult a rational approach to educational development. Nevertheless, over the past two decades most developing countries have engaged in some planning for education—in a number of cases only on a very limited basis.

No two planning agencies follow the same procedures; no two countries' plans are alike. But similar planning steps and methods can be utilized, depending largely on the scope and range of the plan, on the experience of the planners and on the availability of data.

The perspective of educational planning may thus be gained by a description of its scope and process, drawing whenever possible upon experiences in developing countries. To be effective planning generally involves at least three phases: 1) diagnosis and determination of objectives, 2) analysis and program choice, and 3) evaluation of outcomes of implementation. These phases of planning, discussed in the following pages, are not conceived as being unique in their methods since often the same procedures and tools may be applied in the solution of the hierarchy of problems involved.

Diagnosis and the Determination of Objectives

Diagnosis is in effect the first step in the planning process. In this phase of planning the main problems are identified and their scope and significance assessed with a view to specifying national educational objectives and the level of national effort required to achieve those objectives, as well as the broad strategies of educational development in the perspective of time and space.

To be able to define and specify educational objectives adequately, the scope and dimensions of the planning situation must be clarified.

First, the place of education as a basic socioeconomic sector involving benefits and costs should be ascertained. Preparing the future social order in developing countries should not obscure the economic and social role of education at any point in time in the life of a nation. Education absorbs a large portion of limited budgetary resources at the expense of other sectors; it is, however, a sector on which other sectors depend and one which provides employment to a relatively large portion of the population. Furthermore the educational process and educational institutions will plan basic and unending social functions in a nation in all stages of its development.

These dimensions may generally be taken for granted by educators and planners. But the requirements of the educational and planning efforts can be better fulfilled and the chances of success increased if the society at large understands the role and functions of education as well as the great costs involved.

Second, there is need for recognition that the effectiveness of educational planning depends on the degree to which the variety of educational programs are guided, supported, and coordinated. There is need, to borrow Diez-Hochleitner's phrase, for a "balanced approach" in educational planning.

> . . . this balanced approach means that the educational process must be looked upon as a whole. One cannot indulge in the luxury of expanding a given level or aspect of education without considering its effects, the pressure it puts on the other levels, or the restrictions and limitations it imposes on the rest of the process. The whole must be taken into account. The responsibility of the planners, the policy-makers, is not limited to a given aspect of education (19, 67, p. 3).

The educational landscape in developing countries may be varied and complex. In addition to the various types of education, usually under a national ministry of education, other public agencies and private institutions are also directly involved in the business of education and training. In Ethiopia, for instance, the Ministry of Agriculture operates two agricultural institutes at Jimma and Ambo; the Ministry of Public Health runs nursing schools; and the Ministry of Education trains teachers as does the Faculty of Education of Haile Sellassie I University.

The role to be assigned to non-government institutions—such as church, mission and private schools—should be determined. Coördination of the use of nongovernmental resources has been rather difficult in the developing countries because of the lack of information about the role which private or nongovernmental educational institutions can—and will—play in future educational development.

Developing countries can least afford duplication of effort and waste of resources. This can be minimized by a system of planning which is concerned with the entire system and aims at full utilization of resources and capacities in various sectors to meet national educational objectives.

Third, the structure and capability of the planning hierarchy, including the organization and management of educational institutions, leave much to be desired in some developing countries. Often the result is inefficiency and an incapacity to support development and change. On the central level planning responsibility is often delegated to a small technical unit which has neither the personnel nor the jurisdiction to influence decisions of educational strategy and development. Frequently planning units are grafted upon organizations not oriented towards development. Often too, the planning function is neglected entirely on the lower levels and regional centers of development. Under such circumstances it is hardly possible to coordinate effectively the massive educational development envisaged in most developing countries.

Obviously the organization of educational systems—as well as of planning—will vary with constitutional requirements, legislative systems, and the degree of autonomy of local governments. But the actual structure is not significant. More important is the existence of an effective planning system capable of ensuring coordination of the educational effort.

Fourth, it is important to keep in mind that the central feature of planning is rationality in dealing with the future. Since development is a long-term aim, a planner needs to give much attention to temporal aspects. Long-term educational planning can set out national objectives based on estimated economic and social demands and define overall policies and strategies for the attainment of these objectives, including estimates of resources that can be made available for education. The period covered by such an exercise may be 20 or 30 years.

It is understood, of course, that almost everything could change during such a long period and that planners are not so omniscient as to be able to predict all the changes in a cumulative and dynamic process of development. This is why it is necessary to define medium-term plan periods during which more precise targets, resources, and means can be specified.

The period of five years has become a conventional medium-term planning period. The development of the education sector of Ethiopia within the fourth five-year period of national planning was in the main expected to be based on the long-term projections and strategies recommended by the Sector Review Commission. The targets adopted and policies pursued during this plan period would therefore constitute one phase of planning towards the attainment of the long-term aims.

During this time period the educational plan may need revision. Thus a short-term plan, generally one which coincides with the annual budgetary cycle, is required.

The short-term period is critical, for during this time planners often adjust objectives and means in the light of past performances and resources available. Failure to attain objectives specified—by no means a rare occurrence—can confuse priorities unless planners and administrators respond with alternative strategies consistent with both the medium-term and long-term objectives.

The description of planning in terms of the long-term, medium-term and short-term should not obscure the fact that development is a continuous process. Although phasing is required to deal effectively with the cumulative and dynamic nature of development, the degree to which intertemporal coordination is achieved usually leaves much to be desired. This is due partly to the complexity of the undertaking and the limitations of the planning machinery.

Finally, because implementation of plans takes place in relation to given locations as within given periods of time, the spatial dimension in planning deserves much greater attention than actually has been the case. A few of the general considerations pertaining to this dimension will be mentioned here.

The distribution and character of the population and the resources of a country are almost always varied and uneven. This calls for not only the exploration of the possible lines of development in specific territories but also for a distribution of investment depending on the needs and possibilities of these territories.

Existing political and administrative divisions of the country will often set the pattern of distribution or devolution of power as well as of planning responsibility. This is in general desirable from the development viewpoint as the planning effort can be made more fruitful by permitting the design of development strategies geared to meeting the specific needs of particular areas.

Further, it is found that within a developing country the rural and urban sectors show clearly disproportionate development. In Ethiopia, for instance, one estimate puts GDP per capita at E$109 in the rural sector as contrasted with E$680 in the urban sector (Fassil, et al, 29, 1972, p. 65). Distribution of education between the rural and urban sectors is also highly uneven. Different educational policies are therefore indicated under such circumstances.

Last, the problem of location of schools within a given territory often presents itself as a relatively difficult one, as this would have to be determined on the basis of multiple criteria which may include population distribution, availability of land, accessibility, availability of construction material, and so forth.

The basic premise now widely accepted is that education should be related to a broad strategy of human-resource development rather than to a limited concept of "education" planning. Education should be related to national development goals and plans, and the programs of various agencies and organizations.

An essential step is therefore a clear statement of educational objectives based on a diagnosis which explores adequately the dimensions of the planning problem. But this is one of the most difficult tasks facing educational planners in developing countries. National objectives have not yet been formulated in some developing countries, and even where they have been some sound too theoretical. This is one of the first hurdles to be overcome, for unless objectives are stated in measurable terms assessment is very difficult.

Two types of educational objectives, qualitative and quantitative, are needed for the formulation of an educational development strateg . Qualitative objectives state the role that education is expected to play in influencing economic change and in realizing social and political aspirations, while quantitative objectives specify enrollment, financial, economic and related targets. The Ethiopian Education Sector Review has made recommendations on both these levels.

The qualitative objectives recommended by the Conference include the following:

1. to foster a rational and scientific outlook on life; to cultivate objectivity, intellectual curiosity, tolerance and broadmindedness;
2. to replace the traditional negative attitude towards manual work by a positive one;
3. to increase the earning capacity of the individual by providing relevant skills and knowledge; to make people economically self-reliant;
4. to cultivate the desire for life-long education, when formal schooling has been completed;
5. to provide scientific, technical and vocational education, particularly at the secondary level, in keeping with the needs of the Ethiopian society and economy;
6. to Ethiopianize the content of education; to promote the national

language, Amharic, as the medium of instruction at the higher levels, and to give practical orientation to instruction at all levels;

7. to create an integrated society by drawing upon the diverse cultural and linguistic elements and creating the conditions for the formation of a truly national culture;
8. to create national consciousness among all the peoples of the Empire;
9. to foster the full participation of all the peoples of Ethiopia in the life of the nation;
10. to reduce the generation gap between the educated young and the traditionally-oriented old; to bridge the gap between school and society;
11. to prepare the nation's youth to live in a world community;
12. to equalize access to education among all parts of the Empire;
13. to provide universal access to education as rapidly as possible. (23, pp. III-V, 5).

It can be seen that the qualitative objectives of education recommended include reorientation of attitudes and social and political changes, as well as economic betterment of the people.

An example of the quantitative targets recommended is the number of participants on the various levels of the education system projected to the year 2000.

Table 28. Enrollment Targets Recommended by the Ethiopian Education Sector Review

Year	Number of Participants (thousands)			
	First level	Second level	Third level	Nonformal
1979-80	1,988	220	10	838
1989-90	3,071	301	17	1,517
1999-2000	5,273	416	28	2,223

Source: *Report of the Education Sector Review, op. cit.*, Exhibit V-F-1.

These quantitative targets were arrived at on the basis of analysis involving population growth and a strategy emphasizing educational expansion on the first level, and skill-oriented nonformal education for the great number of adult population, particularly in the rural areas.

Quantitative analysis is basic in all phases of planning. For this reason, the Addis Ababa Conference of 1961 called for the establishment of adequate measures for the collection of necessary data and for their periodical evaluation. The types of educational data needed in the various phases of planning are discussed in the next section.

Analysis and Program Choice

Having clarified the dimensions and scope of planning and having specified the qualitative and quantitative objectives, attention can be focused on the more particular and detailed aspects of analysis to serve as the basis of selection of programs which optimally meet the objectives specified. It is on this level of planning that the more rigorous application of planning techniques is required.

The compilation and use of quantitative data is basic for planning analysis. That data, which may be internal or external to the educational sector, may be classified as educational, demographic, financial and economic.

Educational data are basic statistical information about the inputs, processes and outputs of the educational system. These consist of the numbers, distribution and patterns of success and failure of students. They also include the numbers (and level of education) of teachers and other personnel, as well as types and numbers of facilities such as classrooms, books, and so forth.

Demographic data consist of information on the number, composition, distribution and age structure of the population. Such data may often be unavailable in developing countries. The Educational Sector Review Commission in Ethiopia, for instance, was compelled to undertake population projections on the basis of whatever limited data were available in 1972. Population growth in urban and rural sectors as well as by school age was estimated and used as the basis for projection of enrollment to the year 2000. Admittedly such estimates are crude and must be cautiously used. But it is also true that population projections made for such a distant future would always be subject to a significant margin of error.

Financial data may be budgetary estimates or actual expenditures. Such data are usable in planning if available by source for each level and type of education in sufficient detail, distinguishing between salary costs, supplies, library costs, and so forth. Data on capital allocations and expenditures need to be systematically recorded and reported by project and sources of financing. In this respect it should be noted that much investment in developing countries escapes the attention of planners. Self-help and community contributions to education are examples of this kind of unreported expenditure; ways must be found by which such investments can be reported so that planners can make a fuller assessment of educational effort.

Financial data are required by planners for evaluation of efficient use of available resources and for projection of future requirements. For these purposes, it is desirable to develop and make use of indices (for example, unit costs) based on local standards. If reliable indices are developed the problem of projection of requirements of recurrent and capital funds can be simplified. Indeed as the educational sector expands rapidly and becomes complex, evaluation of efficiency and the estimation of financial requirements will become unmanageable without the use of such indices.

Common Techniques of Planning Analysis

The tools of planning analysis vary dependent on the scope and purpose of analysis involved. They may vary from input-output and cost-benefit analysis to firsthand reports by teachers, administrators, and educational inspectors.

Input-output analysis shows the relations of means to ends; cost-benefit analysis attempts to show economic results of proposed policies, and cost-effectiveness analysis compares the costs and gains of proposed policies to help decisionmakers to choose among alternatives. Each has a different purpose. But all are only as good as the outcome measures they employ to capture the specified effectiveness of the strategy, program or development project.

In the growing body of literature on educational planning the use of benefit-cost analysis and manpower studies both receive considerable attention. A brief discussion of the theory and techniques of each method is presented below with examples from studies in developing countries.

The benefit-cost approach is a calculation technique for investment selection. Assuming other things equal an investment with relatively higher yield would be preferred to an investment with relatively lower returns.

For such decisions at least three techniques are used: benefit-cost ratios (usually a ratio of discounted future benefits of an investment to its discounted costs), present net value of investment calculations (where discounted costs are subtracted from the value of discounted benefits), and rate-of-return calculations. Of these the rate-of-return technique is widely used for measuring yield rates from education. The technique is rather simple. To calculate the benefits from a particular type and/or level of education estimates are usually made of the incomes resulting from that type and/or level of education over the working life-time of the average individual. Because a given dollar income in the future is less than the same dollar amount today, the streams of income are discounted.

To date rate-of-return calculations for education are available for many countries (see Table 25, above).

One major weakness of rate-of-return calculations is that educational outputs are defined narrowly by economic or monetary returns.

In the case of less developed countries there are the unique psychological, political and other positive-linked impacts of education which are certainly not accounted for by rate-of-return calculations. The exclusion of all these noneconomic effects of education (referred to in the literature as externalities) from rate-of-return calculations ignores basic justifications for education.

Furthermore expected educational outcomes may not be reflected in income differentials between individuals possessing different types and levels of education. Wage differentials are caused by educational differences as well as family connections, ethnic backgrounds, proximity to work places, and so forth. The adjustment of income differentials for some such factors to isolate the "pure" effects of education in less developed economies is difficult because of lack of data.

Another shortcoming of rate-of-return calculations stems from the common use of cross-section data. The usefulness of such calculations in guiding future educational policies is doubtful. Such calculations may have predictive merits if certain economic differentials remained stable in the future, but such stability is rare in less developed economies where changes occur frequently.

Another problem relates to interpretation of rate-of-return calculations. The calculated single value results call for caution in their interpretations. What are the policy implications, for example, of a lower social rate-of-return in university than secondary education? Some known studies in less developed economies yield such results. In some cases such results are used to support a shift of resources toward levels and types of education with better yields. But the same results may be used for different policy suggestions. For example a recent Planning Commission paper in Ethiopia concludes that the social return to university degrees is so low as to suggest a sharp reduction in expansion rates (**28**, 1973). But an examination of the state of higher education in the country shows clearly the existence of a system which includes no post-graduate training, and with a lumpy fixed-cost situation. Increasing enrollment (and/or more efficient management and use of resources) may reduce costs per student. Greater returns may also be gained by deepening of training and expansion of research activities. The same rate of return calculus may thus be used to advance the case for increasing or curtailing expansion of enrollment on the university level.

The *manpower approach* to educational planning is based on the view that educational outputs must correspond to manpower specifications in national economic plans. Developments in the education sector in many less developed economies have shown clearly the need for such an approach. On the inputs side large proportions of GNP are channeled to finance education. Valuable human resources are also committed to the educational effort. (The average student requires 12 years to go through the process of the educational system in order to arrive at the door-step of "third level" institutions—and this happens in countries where the average life expectancy is not more than 40 years.) Yet despite the clarity with which educational inputs of time, finance, and human resources can be measured, the gains from education are far from obvious. A number of possible educational outcomes are not difficult to see, and some have cast strong doubt on the performance of the education sector in many poor countries. Consider for instance two educational outcomes: wastage and the so-called "brain drain." The large number of unemployed school leavers add a new dimension to the growing urban unemployment problems of most developing nations and there is now increasing evidence that the brain drain is hurting some less developed economies.

The manpower requirements approach in educational planning generally follows the enumerated steps. First, output forecasts are made for a plan period (and usually by economic sectors). From such output forecasts employment forecasts can be made by assuming employment as a function of output (in some stable production function type frame). It is also possible to estimate employment by adjusting output forecasts for changes in labor productivity. A next step is the projection of occupational structure of the sectoral employment forecasts. Finally, occupational forecasts are converted into educational equivalents. The education sector is then entrusted with the task of admitting and providing training for students calculated to meet the level and type of manpower required for the occupational profiles during the projected periods.

The manpower approach of educational planning is also subject to certain criticisms. Although other employment forecasting techniques are available (for example, time series analysis), the usual method is to forecast employment from some employment-output relationship frame. The existence of such a stable production function even in mature economies is questionable. Obviously such a relationship may exist at the firm or factory level but there will be less utility to be derived from relationships of the type when the basis is an economic sector or a whole economy (even when such associations are statistically significant).

It is also equally difficult to determine a stable employment-occupation relationship. As a result, occupational forecasts could be misleading. Some studies indicate a general tendency of industrialization to require a wider diversity of sophisticated skills—for example, a rising proportion of "high level" occupations in each economic sector. Other studies emphasize the sectoral shifts of employment (inherent in the industrialization process) as perhaps a primary cause for changes in the proportions of occupations in the labor force. The rapid shifts in sectoral employment distribution taking place in less developed economies and the diverse technological and management choices open for manning a production unit make the task of occupational forecasts especially difficult in developing economies.

The conversion of occupations into educational equivalents is not an easy task. A primary problem is that for most occupations it is impossible to specify unique educational requirements. It is possible to determine the educational inputs necessary in the case of some occupations such as doctors. But for most activities occupational substitutions are possible, and a given occupation can be carried out by people with different educational backgrounds. Most recent quantitative studies in this area attest to the vague association between occupation and education. These studies reveal difficulties on two levels: on the one hand, the effects of education and occupation on growth and development are unclear and, on the other hand, present evidence suggests the absence of definite association between education and occupation for most occupational categories. Cross-section evidence shows variations between countries in their capacity to employ persons educated beyond lower manpower levels. In any one country it is not hard to observe that a given occupation is associated with varied educational backgrounds.

Experiences in the Application of Analytic Techniques in Developing Countries

Despite the numerous difficulties encountered there is little question that decisionmaking is improved by the pursuit of a logical framework and the use of available techniques and tools in educational planning. The experience in their application and the demand for generating required data can lead to gradual institutionalization of analytically based planning.

Case studies of the use of analysis in educational planning in the developing nations has been undertaken by the International Institute of Educational Planning (summaries in Coombs and Hallak, 14, 1972). Although the application of analytic techniques tends to be *ad hoc*, the studies reveal that use of analytic techniques has assisted planners in solving specific problems in specific situations in a number of developing countries.

For example in preparing the 1966-70 educational plan of Ceylon a cost analysis showed that it would be financially infeasible to increase the primary school participation rate rapidly to 90 percent. A second five-year educational plan of Uganda—to be set within the framework of the overall development plan at the outset—held down primary enrollments so that no expansion of teacher supply was necessary for enrollment targets. Rough costing showed this to exceed the resource limit of the plan, so that another plan aimed at cutting unit costs had to be drawn up. The government, however, considered this to be too minimal to satisfy practical needs and the nation's aspirations. A further alternative plan was then formulated and costed. In Thailand an analysis was made prior to drawing up the formulation of the education sector of the National Economic and Social Development Plan for 1967-71. After targets were set, planners —with UNESCO assistance—were able, in spite of data shortcomings, to test the feasibility and to identify problems and modifications needed in the plan.

The scope and findings of cost-benefit analyses vary. In Brazil for instance, analysis disclosed that the highest social rate of return is from secondary education, the next highest from university education and low returns from primary education (see Table 25). This differs from the findings in Ethiopia cited above where the highest return was found to be in primary education and the lowest from higher education. In India one study attempted to analyze the causes of unemployment among university and secondary school graduates. Two other analyses dealt with the issue of resource allocation and the rate of return on education compared with investment in other capital. The latter studies showed that the rate of return from investment in physical capital was higher than economic returns from secondary and higher education.

The manpower approach in educational planning has been tried in a number of countries, including developing ones. The application of this method, particularly in less developed economies where the data basis for the forecast exercises is usually unsatisfactory, has been criticized by many.

In Ethiopia, for example, there have been attempts to use the manpower approach in educational planning. Criticism of these studies centers on their shaky data basis as well as their methodologies. A first official effort at manpower forecasting was made in connection with the Third Five-Year Development Plan, but the sectoral employment forecasting and the conversion into educational equivalents appear to have been based upon results obtained from Cameroonian and Senegalese experiences. Another high-level manpower forecast failed to show methods used in forecasting (Zack and Mohammed, **70**, 1964). Serious work on manpower projections was done by the Department of Labor of the Ministry of National Community Development and Social Affairs (**22**, 1970). Again the methods employed were strongly criticized and the resultant forecasts consided unacceptable by the Planning Commission, among others. The various manpower forecasting exercises also yielded results marked by extreme variation. The utility of such forecasts for purposes of educational planning has been further weakened by the fact that the forecasts were mainly concerned with "high level" manpower formation.

On a more restricted scope a cost analysis and feasibility study of the proposed program of massive use of television in primary schools was conducted in Ivory Coast with UNESCO assistance (Coombs and Hallak, **14**, 1972). Using crude methods and data, planners were able to complete a study of cost implications of changing secondary school curricula in Madagascar which reportedly has had an impact on government policy.

A fact-finding study in Chile on number and size of classes in secondary technical schools, and on numbers and utilization of teachers, showed the need for better management and use of available resources—not more funds.

The review of educational planning in developing countries leads to the tentative conclusion that the most sophisticated planning process may not lead to plan acceptance and implementation without firm commitment of political leaders and effective cooperation from competent administrators. On the other hand, the crudest processes and plans may be successfully implemented if government leaders support planning, participate in its processes, and view it as essential to the development scheme. Most educational planning in the developing countries finds its place somewhere between these two extremes.

EVALUATION OF EDUCATIONAL OUTCOMES

The necessity for evaluation of educational performance in development planning is hardly a new concern to planners for, ideally, planning is continuous and involves a systematic process of setting objectives, specification and analysis of alternative strategies, as well as evaluation of progress toward objectives. This process is depicted in Diagram 1.

As can be seen from the diagram, the critical link between national development planning and educational planning lies in the articulation of the role that education is expected to play in development and the assessment of educational outcomes in relation with development as well as educational goals.

Developmental goals must be translated into educational objectives and targets. Educational planners can then assess measured outcomes against the specific targets. Where there is inconsistency or infeasibility they must alert other policymakers and/or attempt to initiate reforms in the educational system. The process is thus concerned with evaluation of outcomes, feedback and revision of objectives and means.

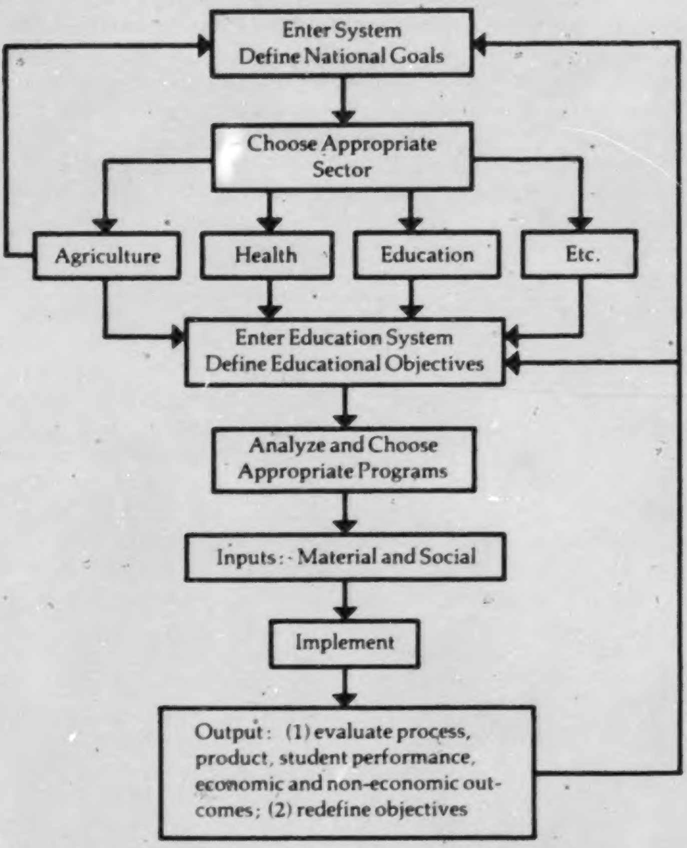
Evaluation differs little from planning analysis in methodology and processes, although there are at least three distinctions. First, analysis is a predecision process of examination while evaluation is an examination after the fact that looks to changes in program and project that can achieve the results more surely or at lower costs. Second, analysis calls for a generation of options to be assessed, and evaluation starts with a given program or project and seeks options only after assessment has been made. Finally, analysis often requires generation of estimates of cost and outcome from a shallow base of experience; whereas, evaluation starts with a body of experience from which data can be extracted.

Evaluation may consist of assessment of the effectiveness of an overall program or of individual projects. Focusing on output rather than input, one of its basic functions is to generate feedback which can provide direction for policymakers for changing current operations, and for planning future programs.

The specific purposes and uses of educational outcomes information are:

1. to assess the capacity, efficiency and effectiveness of present educational institutions and programs in meeting specified objectives;
2. to determine what children and adults, young and old, have actually learned, and to assess alternative ways for improving learning;

Diagram 1. A Schematic Presentation of the Planning Process


Enter System
Define National Goals
Choose Appropriate Sector
Agriculture
Health
Education
Etc.
Enter Education System
Define Educational Objectives
Analyze and Choose Appropriate Programs
Inputs: - Material and Social
Implement
Output: (1) evaluate process, product, student performance, economic and non-economic outcomes; (2) redefine objectives

3. to indicate major failures or weaknesses in the existing programs and institutions;
4. to evaluate and to assist in the design of new programs or policies;
5. to delineate the responsibilities of education and the areas in which intervention by other ministries or agencies would be more effective;
6. to help clarify education's optimal share of the nation's resources;
7. to guide in the allocation of educational manpower and other resources throughout the nation;
8. to help establish a basis for performance for students, teachers, and administrators;
9. to justify requests for new local or external funding when indicated; and

10. to facilitate assessment of planned variations in programs or projects.

As indicated above, education is conceived of as a sub-system within a national development strategy; its function is to contribute toward the fulfillment of a variety of development goals. The ultimate purpose of evaluation is to assess the degree of progress being made toward meeting those goals. This can be achieved in some measure through the evaluation of the educational system itself which consists of particular types of programs designed to meet specified sub-objectives.

Viewed comprehensively evaluation in educational planning must address itself to a hierarchy of outcomes and their measures, such as educational process outcomes, educational product outcomes, student performance outcomes, economic growth outcomes, and social and political outcomes.

Systematic feedback processes are rarely used as components in evaluating the outcomes of a given educational system. Very often the rigorous demands of feedback data systems are not easily fulfilled, but even more limiting is the absence of this type of evaluation system in the minds of most evaluators of education. Furthermore in a developing country, where education has to perform relatively more of the basic catalytic function of social change, the more difficult and less reliable it will be to measure the impact of any educational system by the use of standard evaluation procedures applied in developed countries. As has been indicated earlier, especially when the educational system is borrowed, it will not be quite as easy to assess its impact, particularly the more intangible types of outcomes,

The problems of specifying educational outcome indicators and their measurement are discussed in later chapters. It is sufficient to state here that continual assessment through measurement of outputs is indeed a basic part of its planning process. For it is found that no matter how carefully objectives and strategies are laid out and how many factors and details are considered, it cannot be assumed that matters will develop strictly according to plan; they almost certainly will not. Indeed, by the time new programs are underway the system may look very different from its blueprint. The administrator or the person in charge of implementing the plan needs to bear in mind that an educational plan must constantly be amended and sometimes overhauled. Plans are implemented essentially where the job is done.

12

Use of Outcome Measures in Schools

Teachers are the vital link between the administration and the students. They are the interpreters of the program and its goals. They can, in fact, greatly affect whether the objectives will be achieved. They are also the first to observe changes in knowledge, behavior and attitudes; they have much to gain from a systematic way of utilizing that information to assess students and their own teaching practices. Therefore they need to be exposed to the purposes of the program, and to become familiar with the standards or outcome measures used to determine achievement of those purposes.

TEACHER TRAINING

It would be opportune during the period of teacher training either in-service or pre-service, to introduce teachers and student teachers to the specific purposes of the programs in which they will participate and the criteria by which those programs are being evaluated. A suggested orientation for teachers might include the following topics: why it is important to measure outcome; the processes of defining specific learning goals to be measured, including teacher participation in goal selection; and actual preparation of instruments to measure goal attainment (for example, reliable and valid tests, good observational rating scales, interviewer questionnaires); and interpretation of outcome measurements.

Incorporation of this knowledge in training programs for teachers would help to keep teachers focused on the actual relevant outcomes of the educational process. Where outcomes are falling short of targets, teachers may be in the best position to determine why. For example, in an adult literacy class attendance may be good, interest apparently high and results on the class quizzes adequate, but beyond the classroom no one is using the skill. Further exploration may indicate that the only available reading material is highly technical and the language of the primer is considered too childish to be useful. A creative and alert teacher might sense the need to change the text or write a new one, to help people write letters in class, or to encourage a collective effort to secure a modest library in order to get the program to produce the desired results of functional literacy.

Existing Measures: Achievement Tests

Most teachers encounter some form of externally prepared standardized exam. Thus they should learn: who prepares it, who uses it, for what purposes; how the test is given, graded, interpreted; and how the test has been or can be revised.

This type of training can also familiarize the teacher with how to use the test. For example, teachers can ask:

1. Do I teach the information covered on the test? In other words, can and do I use the test as a guide to teaching?
2. Does the test cover the information I teach? The answer to this may lead the teacher to communicate with test designers to encourage them to incorporate content covered in class. The teachers may even be drawn into the process of test design themselves.
3. Is there a pattern of student success or failure shown by test results which I can correct? The pattern may indicate the need to change course emphasis, to spend more time with slow learners, to attempt to approach the material using a different method of teaching.

Teacher training in outcome measurement for diagnostic purposes should not neglect the area of teacher attitudes, expectations and interests and its impact on accurate student or program assessment. An awareness by teachers of their own and their students' nonintellectual traits and needs may well in itself change the climate of the classroom to foster learning and development. Many studies on the effects of teachers' expectations for students' achievement have demonstrated that students are influenced by the opinions of their teachers. Indeed some researchers feel that teacher attitudes are the missing element in studies such as the Coleman Report in the United States and the International Association for the Evaluation of Education (IEA) assessment, which seek to establish the variables associated with student achievement.

Training can also guide teachers to develop their own evaluating acumen. This, of course, is one of the purposes of teacher training: to show teachers how to measure the effectiveness of their teaching. Introducing the trainees to the principles and particular methods of evaluation should be relevant to that purpose. Formal preservice or lengthy in-service sabbatical training provide the opportunity to explore the theory, practice, and history of the development of outcomes assessment. At this time teachers may be offered the opportunity to participate in the design or reconstruction of nationally used standardized tests. They may be trained to develop a whole program of evaluation and assessment for their region (as was the case in Indonesia), for their school, their department or program. Training to use outcome measurements should stress the measures the teachers actually encounter or those they could develop, such as in-class tests appropriate for particular circumstances.

Hopefully, teachers trained in output measurement will become better teachers by having at their disposal a systematic basis to assess and revise their teaching practices in light of student performance. Another point that supports the argument for such specific training is that getting teachers involved can help improve the accuracy, applicability and quality of assessment measures and the entire planning process, for too often the evaluation of students and programs has been an alien process conducted by administrators or officials remote from the actual learning conditions.

ADMINISTRATION

To fulfill their responsibilities, administrators need effective methods of assessing needs and the efficiency of programs. Because of the pivotal position of administrators in the educational hierarchy it is here that much of the outcome data can prove most useful.

A major concern of administration is the efficient allocation of the resources between teacher and staff salaries, materials and plant to ensure high productivity and low costs.

For this assessment they can use analytical techniques employing investment measures such as cost-benefit, cost-effectiveness analysis. These are quantitative measures which can be used to relate personnel and material costs to such factors as learning, graduates, jobs, attributes, and attitudes.

With regard to teacher salaries and promotion administrators can examine their incentives and workloads in the light of teacher satisfaction and turnover, and how this affects student outcomes. They may ask: Are better teachers promoted out of the teaching role into administration? Are teachers paid according to actual teaching ability and student performance, prior educational credentials, or seniority?

They may select, fire, promote, and arrange the schedules of teachers. To assist them in this process and in relating student performance to the influence of each of these internal components of the educational system, administrators can utilize the student output measures.

For example, administrators in schools where performance has been low must ask themselves if curricular manipulation, better trained teachers, or more efficient tracking patterns for the students would change their outcomes, or if the probable causes of below average performance lie elsewhere. (It has been shown that health and nutrition are important influences on student performance, capacity, and interest. Perhaps the introduction of a school breakfast or food supplement program would in fact improve learning.)

Furthermore administrators need tools to evaluate student academic performance and behavior. Results from national or regional external examinations can be used to compare students in similar programs, to determine how many and which types of students are progressing at an average or "normal" rate. Where external tests are not available or not applicable, administrators may have to commission university researchers and psychometricians or their own staff to design new measures. Such data are valuable in evaluating the effectiveness of the school.

The administrator is at the juncture in the educational hierarchy where accountability, responsibility and authority for many decisions coincide. The focus that the administration establishes, the kinds of system objectives pursued, and the measures used to gauge the achievement of these objectives will not only affect the efficacy of the management of the system but will serve to guide teachers, parents, and students.

Administrators therefore must consider themselves and their procedures as variables in the educational milieu. Administrative procedures often pervade every aspect of the educational environment. They determine class size, length of class time and school day, disciplinary measures, location of schools, use of educational plant when school is not in progress, provision of ancillary services such as guidance, nutrition and health care, libraries, tuition and other costs to students and many other areas. These procedures must be scrutinized to weigh their effects on efficiency and student learning, and to accomplish this a greater use of outcome measures would be helpful.

CURRICULUM AND METHODS

A variety of outcome measures and methods for evaluating education were discussed in previous chapters. Now we turn to how such measures can affect the internal aspects of education. Specifically, we examine the use of those outcome measures for curriculum and teaching design.

A concomitant of course selection and design is the development and production of instructional materials such as blackboards, pencils, paper, notebooks, tape recorders, projectors, texts, records, films, television or radio programs, and laboratories. But without detailed knowledge it may be difficult to understand the effects of those different instructional methods.

In designing courses outcome data can provide answers to the many problems confronting planners and administrators. Such data can shed light on the appropriate sequence of courses; necessary variations in courses for different age groups, regions or cultures; and what level of difficulty is appropriate for a given student body.

Outcome measurement also can inform teachers about the influences of various teaching methods such as the lecture, group recitation, group discussion, laboratory or practice, and project methods. Each method varies in regard to student-to-student interaction; student-teacher interaction and feedback; and opportunity for self-discovery, student participation, individualized pacing, and discipline. And since there are such variations it can be expected to follow that one method might be more effective in a given course or student situation; outcome data help to pinpoint these areas.

When specific outcome data are applied, the major questions in methodology that can be addressed include:

- Which methods are most suitable for each type of learner, for example, the slow learner, the older child, the adult, the very young?
- Which methods are most suited to each of the various learning environments?
- Which methods convey certain subject matter better than others?

The output measures mentioned in this volume can be used to explore the following issues involved in curriculum formulation:

- Are students ready to handle the material presented; do they have the proper foundation in skills (reading, for example) and information.
- Is material presented in a logical, orderly fashion.
- Are the methods of presentation suitable to individual differences in ability and consistent with regional, national or tribal cognitive styles.
- Does the method encourage memorization of facts or comprehension of principles.
- Is the subject matter suitable to expected conditions of employment and living.
- Do the materials used, such as laboratory equipment or farm tools, facilitate learning.

One of the more widely used devices for measuring educational outcomes is the standardized test. These tests can be used in several different ways to improve the internal aspects of education.

Here is one example: Matched groups of students are taught some course in which each group is subject to a different curricular variable—lecture vs. discussion method, T.V. program vs. live teacher only—or an emphasis on inductive vs. deductive reasoning processes. A standardized test administered at the conclusion of the course may demonstrate better overall achievement on the part of one group. With careful analysis these results may indicate which of the two curricular approaches is better, and in which ways.

And standardized tests, by measuring student performance, may reveal weak links in courses. They may indicate that students of different regions, cultures, family backgrounds, sexes or aptitudes need differing kinds of treatment. This could alert teachers and administrators to specialized student needs and lead to tracking of students into different classes. Furthermore planners, curriculum designers and teachers could restructure courses to compensate for the test-indicated lacks or to reemphasize and utilize these approaches which gave good results.

Another example of how outcome measures such as standardized tests can improve the internal aspects of education is in course composition itself. Recent research has demonstrated the value of breaking down a course into modules, having a specific set of learning objectives for each unit, and using a series of tests to measure the desired objectives. If there is a failure to progress to the next module in a specified time then it falls to the teacher and to the school to determine what went wrong and find solutions. This procedure is valuable since it can be applied immediately as a teacher guide to overcome deficiencies.

Where standardized tests are not available or inadequate, teachers can be trained to test or measure student performance, objectively or subjectively, and to attempt to relate the outcomes to their own methods, the course content, timing, and so forth.

More than periodic testing may be necessary to provide a stimulus for curriculum reexamination and reform. Among other things the extent of mastery of farming, carpentry, typing, and other vocational subjects can be assessed only by the quantity and quality of output and proficiency. Work skill measurement can be incorporated into the academic course as well. A course in literacy, for example, may require the writing of a short story by which knowledge of the tools and ability to employ them is demonstrated. Of course such testing can be feedback to the teacher or curriculum designer for planning new courses or modifying existing ones.

The measures discussed so far are all short-term performance indicators which demonstrate interest and mastery while in school. Since the purpose of education is to provide some useful way of looking at the world and some useful skills to operate in it, no evaluation could be complete without feedback, latter-stage school performance, employment, and life application.

For example subsequent student performance in school, culled from test scores, teacher ratings or student selection of a course of study, can be part of a curriculum planner's data base. In England educators were curious about student college preference for the pure sciences rather than engineering. The reasons were traced back to training in the 6th Form where course emphasis was on theoretical issues (Maclure, **44**, 1968).

In the United States in the 1930s a massive study was conducted of the graduates of 30 high schools that had experimental programs. 2,108 graduates of these schools were matched with high school graduates from traditional high school programs and the progress of each group in college was followed. Data came from student interviews, questionnaires, reports from instructors, official college records and comments from others who had contact with the students. Differences between the two groups in cognitive achievement, effectiveness in solving problems of adjustment, concern for world affairs and other areas were significant, favoring the students from the experimental programs. Educators planning for these goals could have used these results as a reflection of instructional methodologies (Gage, **32**, 1963).

Planners may also want to design and evaluate curricula on the basis of what opportunity the program provides the student to re-enter the educational stream at a later date.

If this concept of a recurrent education is chosen as a goal it will require a number of significant changes. Planners and administrators in the nonformal and vocational programs will have to redesign some parts of their courses so they adequately prepare students who desire it to shift into the formal system. Administrators in the formal system will have to reevaluate their entrance requirements, relying more on knowledge and skill than on years of formal schooling, to be able to accommodate those whose education has occurred in a different milieu.

Employment patterns can be studied and employer feedback solicited. In a labor market where a range of job options is available student job selection may reflect on the employment preferences instilled by educational training. If students will not take anything but white collar jobs, as is often the problem in developing countries, then planners can look at the type of students this attitude is prevalent among and return to the programs they participated in to see if there is not some change in curriculum possible to discourage this attitude.

Also, transferability of skills can be explored. Is the educational program providing a broad enough basis for its students to find and perform work in a changing world? In Chile, follow-up studies were done on graduates from vocational and academic (liceo) programs. After six months on the job the academic students were more efficient than the students specifically trained for the job. Coordination with employer needs and reexamination of the methods of teaching in the vocational program could lead to curricular reform (Schiefelbein, **60**, 1971).

Attempts can be made to assess direct utilization of skills or information. In Ethiopia the WOALP Rural Development Program interviewed farmers after the project's completion to see if they were using techniques that had been taught. Different farmers had been involved in different programs of various duration and intensity. In this instance good comparative data was collected on the performance of each group and then fed back to the program designers to select the most effective treatment.

Such follow-up measuring could also be used in literacy and health practices. The best measure of literacy is the degree to which the new literate reads available material, visits libraries, writes letters or serves as village scribe. And a long-term outcome measure of the success of a health care program would be changes in infant mortality, or the incidence of epidemics and illness in an area, or utilization of available medical facilities.

"THE HIDDEN CURRICULUM"

Another ingredient in the educational process is, "How has schooling or learning affected a student's motivation, attitudes and emotions? How has it shaped the person?" Interest in this has increased as people recognize that schools condition individuals as well as produce trained manpower.

When certain aspects of school organization are not recognized as having impacts on students, as is frequently the case, they become part of the "hidden curriculum"—educational content which is not planned or organized and which results in student outcomes which are not anticipated, acknowledged, or evaluated.

The school system has modernizing effects on students. A concern in some developing countries is that the changes occurring in the course of modernization and urbanization have been accompanied by alienation and psychological malaise. It is difficult if not impossible to establish a clear cause and effect relationship between these developments, but it can be said that rapid change in lifestyle has psychological consequences.

Many observers have noted that school attendance, regardless of curriculum content, has a modernizing influence on students; some have wondered if it does not also produce a sense of alienation. Schools differ, but many if not most are characterized by large groups of children sitting still for long periods of time, listening, repeating, and memorizing. They become accustomed to routine, mental rather than physical activity, and verbal as opposed to practical experience. The hierarchical structure of school personnel, expectations in dress and in hygiene, division into age groups, and depersonalized relationships with teachers may be contrary to home experience and foster a sense of regimentation.

There are several methods of measuring these noncognitive outcomes. They may also be looked at from the perspective of which treatments contributed to particular attitudes.

In Chapter 3 there is a discussion of a range of scales and testing procedures which assess a person's feelings of self-control, attitudes toward one's self and others, toward school and learning in general, and motivation to achieve. Careful application of these measures during a period of instruction can help reveal student reaction to the schooling situation itself.

Where specific tests cannot be administered, observation by a trained observer may prove effective. There have been a number of attempts to develop categorical check lists and rating scales which can aid an onlooker in his perception and judgment of a classroom situation. There are always the problems of observer bias, the disruptive effect of an outsider's presence, and variability in teacher or student behavior from one day to the next. Despite these drawbacks observation is a direct way to get information on whether students look happy as they play or what in fact they do when the teacher leaves the room.

Direct questioning of students about their feelings or attitudes toward some particular curriculum can also be a useful measure—as well as provide them with the opportunity to participate in the evaluation process.

Educational psychologists have found that there is no hard and fast line of demarcation that can be drawn between cognitive and affective skills. Subjects such as geography and history clearly are not simply an intellectual exercise if the version taught in school conflicts with local traditions.

Even in a relatively "pure" cognitive area such as science, attitude plays a major role in the learning process. The teaching of science illustrates the ways in which the affective characteristics of children interact with the cognitive subject matter. The first question to be asked in planning the curriculum is, "Does this material conflict with the value and belief systems of the students?" An American teacher in Nigeria once described the disbelief and dismay of college students on hearing that he had loved chemistry as a youth because the laws and reactions were so regular and predictable. The Nigerian students had always believed that the natural world was unknowable and uncontrollable (Horton, 34, 1967). In this kind of situation the contrary scientific or rational point of view should be presented with tact. The teacher must determine whether the attitude of the students toward logic and the scientific methods is favorable, hostile, or open; whether derived from the home environment or from earlier schooling experiences, these attitudes affect future learning. The sophisticated rules of logic and other features of modern scientific method have evolved over generations. To some degree appreciation of scientific method in developing societies may require experiencing an accelerated recapitulation of that long history of the scientific inquiry. Patience and awareness of the need for relevance to the local society are important attributes to the science teacher. A third aspect of science instruction is whether it tends to place a higher value on "pure" versus "applied" science—which has long-term vocational consequences for students. A fourth important point is that teachers need to be aware of the possibility of inculcating discriminatory conceptions of ability. As indicated earlier, in many cultures girls do poorly in science and mathematics. These subjects have been defined as "masculine" activities and the girls (like most people) generally perform in line with others' expectations.

Other school functions such as the creative arts, civic education, sports, physical education, health education, adult continuing education and preschool programs all have direct noncognitive outcomes.

Consider the case of preprimary programs. Many children enter the first grade already suffering from malnutrition and inadequate social, psychological and physical care. Since physical neglect and malnutrition can retard the learning of a child for life—malnutrition was recently identified as a major factor in achievement by the IEA—such preprimary programs are needed in depressed urban and rural areas.

Other such valuable programs might be implemented within the school structure which on first glance would seem to have no relation to the task of imparting knowledge. However, with creative use of outcome measures the clues to improving a child's or an adult's capacity to learn might be discovered and innovative programs designed.

In these areas as in other components of the "hidden curriculum," the use of outcome measures can provide valuable insight into the nonacademic products and processes of an educational experience. Feedback from the results of such evaluation can aid curriculum planners in designing learning situations which foster desired student behaviors and attitudes.

13

Improving Resource Allocation

The basic elements of educational and development planning were introduced and explained in Chapter 3. In subsequent chapters educational output measures have been introduced and examined in a variety of forms suggested by the distinct goals of development. The task of this chapter is to reintroduce the planning framework and suggest how outcome measures are a vital component of the planning process. This is a policy-oriented chapter designed to make explicit the practicality of the results of previous chapters where output measures were developed. Primary emphasis will be given to systems analysis as a planning tool and, in particular, the system in which analysis takes place—namely, a planning-programming-budgeting system.

The need for outcome measures in a planning framework is obvious—without some idea as to what is actually produced there is little hope of providing for future needs. Examples of the use of outcome measures in planning are discussed in this chapter.

The basic planning techniques introduced earlier are manpower requirements planning and the rate-of-return approach. Although not specific to any particular development objective they are most closely identified with planning for economic objectives.

The manpower approach is based on a static view of education for development that emphasizes the outputs of education in relation to the needs for those outputs. Surveying the education sector, the planners would try to answer the question: what are the likely changes in the numbers of persons with specific occupational skills? Using this information in conjunction with the numbers of individuals with differing skills, the expected changes due to death, retirement, etc., it is possible to estimate the manpower supply. These estimates, used in conjunction with some projections of demand, allow the planner to judge the probable educational expansion necessary to reconcile demand and supply.

An important ingredient in this approach is the assessment of the educational system's production of needed manpower skills. Thus, for accurate manpower projections outcome measures are needed to assess those skills. For example literacy tests, work sample tests, supervisor's reports and other outcome measures may provide documentation without which any manpower projection would be inaccurate and misleading.

Rate-of-return analysis, like manpower planning, is a planning tool aimed at economic objectives. Again the advantages and problems with this approach have been examined previously. The essence of the approach lies in using estimates of the benefits and costs of various forms of education to form rates-of-return over time. Planning decisions are made by comparing rates-of-return for various proposed education investments.

Rates-of-return purport to gauge the productivity-increasing effects of education; however, due to problems of interpretation and conception they too may be inaccurate and misleading. In this case the planning tool and the output measure coincide: rates-of-return are guides to resource allocation and an estimate of the investment effect of education.

THE APPROACH OF SYSTEMS ANALYSIS

The approach of systems analysis provides a logical framework of planning analysis. This approach requires that the problems involved are adequately explored and evaluated.

Systems analysis essentially involves a reduction of complex problems into their component segments so that each segment can be studied separately. Questions of fact can be subjected to scrutiny through this process while aspects of the problems involving judgments can be identified and the basis of judgment made explicit.

The basic elements of systems analysis are the following:

- a clear definition of the problem(s);
- identification of specific objectives involved;
- selection of criteria or measures of "effectiveness" which will permit estimation of progress toward the objectives;
- identification and description of the key features of the alternative ways of attempting to solve the problem(s);
- estimating the full-cost implications of each alternative including future as well as immediate implications;
- a clear presentation of the "tradeoffs" among the alternatives considering the costs and effects;
- identification and quantification of the major uncertainties, and, if possible, estimation of the effects of these uncertainties on potential decisions;
- identification of the major assumptions made in the study with an indication of the degree to which program choices may be sensitive to these assumptions; and
- documentation to permit others to understand and evaluate what was done in the analysis and to obtain a judgment as to the accuracy of the basic data and the findings of analysis. (Mushkin and Pollack, 51,) 1970, pp. 330-331).

Although systems analysis may use, if applicable, many of the techniques of mathematics, operations research and economics, it may require no more technical sophistication than the pulling together of already existing data in a meaningful and informative way.

Analytical tools are applied continuously with varying degrees of intensity or rigor and scope of inquiry. Furthermore it is possible to generalize and adapt a system model to varying situations, as Razik demonstrated when he applied the techniques to educational planning in Sudan, Ceylon and Kuwait (56, 1972).

THE PPB SYSTEM PERSPECTIVE

The PPB system involves a formulation of the planning problem focusing on a structuring of program and on the linkages between program formulation and budget allocations. The focal element is analysis of program alternatives.

The outstanding features of PPB systems briefly described below pertain to the design and structure of programs for priority setting, and the use of its method of program budgeting.

Once agreement is reached on objectives, commitment to those objectives can be demonstrated in part by how well they are translated into actual programs and budgets. The programs specify key elements for particular analysis and decisionmaking. Specific techniques and tools of analysis, often used in systems analysis can be applied and, based on such analysis alternatives can be quantified in terms of the specified objectives.

The types of program options that may be analyzed are exemplified by those presented to the Education Sector Review Conference in Ethiopia. A central problem faced by the Conference pertained to the selection of programs of primary, secondary and nonformal education which would optimally meet the objectives it adopted. The three options presented to the Conference were:

1. continuation of the existing programs, consisting of six years primary, two years junior secondary, four years senior secondary, and a relatively limited nonformal education program;
2. four years of minimum formation, four years of second level, and four years of senior secondary level education, and a nonformal program with enrollment of about 14 times that of alternative (1); and
3. four years of minimum formation education, two years of junior secondary education for selected students, four years of senior secondary education for some students, along with two to three years of basic formation education for older students followed by three years secondary formation education for some.

The alternative options were presented with data on projections of enrollments, ages of students in the various levels, school days per year, shifts per day, pupils per class, teacher requirements, and costs. The Conference debated the alternative options in terms of these factors and the objectives already established. Interestingly and as is usually the case, it chose to combine alternatives presented and recommended a fourth option which combined elements of two of the alternatives proposed.

In such a setting of decisionmaking the PPB approach can lead to a sharper presentation of the issues, more formalized analysis and better assessment of the effectiveness of the alternatives in terms of the demands of objectives and the critical problems involved. Hence, the basis of decisionmaking could be greatly improved.

Program structures in a PPB system become the basis for budgetary allocations, hence the name "program budgeting." Traditionally, budgets are assigned to schools according to more or less standard types of expenditure categories such as salaries, equipment, travel, etc. The program budgeting approach allocates budgets to defined program elements according to purposes and outputs. The value of such a procedure even for defining total cost of a program can be appreciated from the following example.

Assume, for instance, that a program is formulated for the development of a Faculty of Medicine in a given university. If program budgeting were to be applied the costs of the program could be broken down into sub-program categories: research, patient care, and teaching. Ordinarily only direct costs classified in standard accounting categories would be charged to that faculty. The true cost of the teaching program could be reflected by charging to it, for instance, the cost of teaching done by other faculties to students of that faculty and other costs of the overhead type. This permits the development of a valid cost per student for that faculty which is valuable in educational planning.

Needless to say difficulties arise in the attempt to implement a PPB system. But some of the difficulties that are encountered would largely amount to bringing to the surface fundamental issues which should in any case be faced head on if serious planning were desired. Admittedly the difficulties related to the quantitative analysis that would have to be made as the basis for the selection of programs would not be easy to surmount because of the limitations of data as well as technical capability. But if programs were analytically linked with objectives and program budgeting were achieved, there is little doubt that benefits would accrue, particularly in the greater amount of program information that comes to be available.

Is an adoption of only some of the elements of PPBS possible? Some proponents of the system answer positively. A recent recommendation for the adoption of program budgeting by the Ministry of Education of Ethiopia is presented as follows:

> It is recommended that the Ministry of Education and Fine Arts begin an unsophisticated form of programme budget. The programme

budget would provide information for decision-making. It would be compatible with the present accounting/budgeting procedure, and would supplement, not replace, the present system. Some of the details of the programme budget are suggested, but it is recommended that the operating departments modify these in light of their needs and in light of the information available (Ethiopia, Planning and Programming Office, Ministry of Education and Fine Arts, 27, 1972).

ELEMENTS OF PPB SYSTEMS

PPB systems are variously defined and adapted by governments. We define PPB here to mean a system of inquiry about, and management of, public programs and activities by objectives. It is governmental programming by objectives. The formulation and assessment of those objectives, examination of the alternative programs that can achieve them, measurement of resource requirements, and accountability for program results are fundamental to the system.

We view PPB in the context of a system—a system for bringing together informative documents that as a routine process of management can provide policy officials with more and better information. The information that would be provided includes carefully examined purposes, program costs, and potential program results in achieving specified purposes through various program options, both immediately and in the longer run. The routine of the system requires analytical documentation prior to official budget and program recommendations. And as a system PPB requires the orderly processing of analytical work so that the timing is appropriate for the cycle of work in budget preparation.

It is a system for 1) helping to achieve management by objectives 2) formulating programs in relation to operationally defined objectives 3) generating new program designs and searching for alternatives, 4) assembling total program costs, and 5) analyzing those programs in accord with specified criteria for measurement as a basis for resource decisions.

By requiring a search for options, PPB opens the door to invention and innovation.

It provides an occasion, too, for the consideration of interacting and interrelated activities that serve common purposes. That is, it sets the stage for interagency dialogue, communication and more efficient coordination. For example, many government ministries are concerned with educational achievement; education cannot be considered as an interest exclusive to the education ministry.

Elements of the PPB process generally are not new, but their combination and systematic application to education and other public affairs is. A series of documents has been defined to constitute the tools through which the PPB system is implemented. In addition to evaluation studies, they are:

1) the program structure and statement of objectives;

2) program analyses (cost-effectiveness analyses) and memoranda; and
3) the multi-year program and financial plan.

The components of the PPB system, the documentary tools, and the processes are outlined in Table 29.

Preparation of the several documents of a PPB system requires:

1) clarifying and specifying the ultimate goals or objectives of each activity for which a government budgets money;
2) gathering contributing activities into comprehensive categories or programs to achieve the specified objectives;
3) examining as a continuous process how well each activity or program has done—its effectiveness—as a first step toward improving or even eliminating it;
4) analyzing proposed improvements or new program proposals to see how effective they may be in achieving program goals;
5) projecting the entire costs of each proposal not only for the first year but for several subsequent years; and
6) formulating a plan, based in part on the analysis of program cost and effectiveness, that leads to implementation through the budget.

The statements of objectives of governments, and of education as a function of governments, are fundamental to program structuring, analysis, long-range planning, and evaluation. Unless it is clear what outputs are being sought there is essentially no way of knowing whether agencies are achieving them.

Formulation and definition of purposes require that cabinet officers and ministry heads ask anew about their goals in public service. What is it that is being sought by way of results or products? Or what needs doing and for whom? Formulation also compels an inquiry into the following areas: Why is each current activity being performed? For example what are the purposes of human resources programs? Is the main objective to raise the level of output in the nation—to increase, that is, the nation's productivity and gross national product? Or is it to improve the level of living and per capita income of the poorest in the nation? Or to develop the intellectual capacity of the nation for cultural pursuits? Or is it to facilitate introduction of electricity in rural areas by training linesmen?

These purposes differ and so necessarily would the criteria by which progress could be judged. The type of program that is designed would also differ depending upon the choice of purpose made. If all these purposes are sought each still needs to be identified and measured separately. Those who must decide can judge in what combination they would emphasize national economic growth, amelioration of poverty, or cultural development, for example.

Central to the PPB effort is analysis of the initial objectives and probable results and costs of programs.

Table 29. Components, Tools and Processes of a PPB System

Component Elements	Documentary Tools	Processes Required
Structural	Statement of Objectives	• Formulating and defining objectives • Formulating criteria of measurement
	Program structure	• Classifying programs and activities into a hierarchy • Assigning expenditures to classification of program categories and elements
	Multi-year program and financial plan	• Summarizing decisions taken in output and cost terms • Projecting program levels ahead • Projecting workload costs of current program ahead
Analytical	Program analysis study (problem definition statement)	• Defining objectives • Defining criteria of measurement • Formulating program options
	(Cost-effectiveness studies)	• Developing model for analysis • Collecting data relevant to criteria of effectiveness • Collecting relevant cost data • Carrying out data analysis
Evaluative	Program evaluation studies	• Collecting data on program performance • Designing experiments where indicated

Planning, programming, budgeting systems, as systems, are best applied on a government-wide basis in which defined objectives apply to the entire range of governmental activity. (Government-wide is being defined here to mean an independent taxing-spending decision unit.) The government-wide effort is an attempt to gain better understanding of the range of programs and of ministries concerned with the same or similar objectives. It attempts to provide a process within which the ministry of education, for example, may see the scope of current services under their direction that is important to satisfying the goals of other ministries or of the government as a whole. Similarly, the ministry of education within this process may better comprehend the contribution that noneducational ministries make to the mission of intellectual development of a population. When economic development is a central government purpose, the comprehensiveness of approaches to structuring the PPB system's work and analysis becomes critical.

Comprehensiveness in goal setting and structuring makes program analysis the starting point of the PPB process, especially in the developing nations of the world.

The analysis process is a unifying and comparing one. On the one hand program consequences are assessed in costs both immediate and subsequent and, on the other hand, they are assessed in benefits or effectiveness. Furthermore these factors are considered side by side for various program alternatives to make rational decisions more likely. Analysis essentially involves a reduction of complex problems into their component segments for closer scrutiny.

Recommendations made on the basis of analysis within the procedures of a PPB system are presented in policy papers termed "program memoranda." The "program memo" is a document covering one major program area or a major portion of a major program area. Its purpose is to present major program policy findings, specific recommendations, and the reasons for these recommendations, including a summary of analyses. It is submitted prior to detailed budget preparation.

In general hundreds on hundreds of program problems and issues would lend themselves to detailed analysis within each department or government, and many thousands for a nation. The number of problems far exceeds the analytical staff resources even in a nation as well endowed as the United States. In the developing countries personnel limitations are severe. It is not possible, therefore, to analyze each issue in detail at the outset, or even over several years. It becomes necessary to choose a few issues for detailed analytical study. And for some issues so selected, the period of study required may postpone its immediate use for policy decision.

EXAMPLE OF OUTPUT MEASUREMENT IN A SYSTEMS APPROACH

The use of educational outcome measures in a systems approach can be illustrated by an example of an issue analysis—illiteracy. High rates of illiteracy are a common problem that have many implications for developing countries. Illiteracy may impede increases in agricultural or manufacturing productivity, inhibit social or political change, and hinder the spread of family planning or health information. Within the broader context of development planning illiteracy can pose a barrier to the accomplishment of any of the broad range of development objectives discussed previously. In what follows a step-by-step, yet brief, example of the questioning processes of a PPB system will be laid out to demonstrate the technique and its application to an important problem.

Step one is to state the issue confronting policymakers. That issue may be stated simply as how best to improve literacy among all segments of the population. Notice that before even this first step is taken it has been determined, by whatever process is used, that illiteracy is a problem for the nation. The basic issue or issues may be stated more concretely than simply how best to improve literacy rates. It may be that there are specific programs currently under public scrutiny, and the problem is which to employ.

Step two is to define the problem. Broadly, the problem is illiteracy (to be defined) which may range as high as 95 percent in some less developed countries. But the problem must be examined more closely if particular solutions are to be proposed and either accepted or rejected. For example a further question is what are the effects of the problem? In the case of illiteracy its effects reduce productivity, social mobility, political participation, and communication of problems. The cause of the problem is also important but in this case only vague reasons may be cited—such as fundamental and deep poverty (cause and effect may be intertwined here). Another facet of interest is determining whether the problem is uniformly spread among all segments of the population or isolated in particular social or economic groups. Obviously such a question is important in proposing alternative solutions. For example questions include the breakdown of the extent of illiteracy between rural and urban persons, ethnic groups, geographical areas, religious groups, and so forth. (As indicated earlier, UNESCO data show that there is a major discrepancy between illiteracy rates for urban and rural citizens.) In summary, a clear and complete statement of the problem is required before any decisions can be made.

Step three is to state basic objectives explicitly. So far the objective has been general. Yet the question can logically be posed—improve literacy rates for what reason? (Some of these basic reasons have been mentioned previously, for example, to improve agricultural productivity.) Depending upon what the basic objectives are, the policy solutions may differ and hence a clear statement of objectives is required.

Step four is to select criteria that can assess progress toward meeting the specified objectives. As such, this is the step prior to proposing and examining alternatives. In short, specify the grounds upon which major alternative solutions will be judged, making sure the criteria are complete, specific, and measurable.

At this point output measures become important and must be explicitly introduced, for accomplishment of objectives can be gauged only by proper measurement of outcomes. For instance, for a broad objective of increasing literacy from 40 to 60 percent, the output measure and the accompanying criteria might be performance, by school leavers, on a criterion-referenced standardized test. Alternatively, if the rural productivity objective of literacy training is specified, then measures of productivity and farm yield are required. Criteria, outcomes, and objectives are tied together in the planning process.

Step five is to summarize the major alternatives. Alternatives in combating illiteracy include increased use of existing educational facilities by enlarging the scale and coverage, adult literacy programs, programs aimed at rural poverty and literacy such as Work Oriented Adult Literacy (WOALP) and mobile units for short-term instruction. All of these programs will reduce illiteracy according to simple criterion-referenced tests; some will also work toward other objectives as well—such as direct productivity consequences of instruction; and finally, the time factor and coverage will differ among the alternatives.

Step six is to estimate the costs of each alternative. Analysts have to determine which processes of the alternatives give rise to costs, the categories of costs (fixed or variable), and the indirect cost consequences of the program both for individuals and society. Much of this technical information may be well-developed and available except for innovative programs.

Step seven is to assess the effectiveness of each alternative to examine both costs and efficiencies. This step ties in with the criteria adopted and the objectives specified to reach some conclusions about the efficacy of the programs. Each option has to be examined for its probable consequences and the timing and coverage of those consequences detailed. Moreover the assumptions behind conclusions must be made explicit. If simple, standard literacy grades are the criteria of effectiveness, then the probable impact on nationwide literacy of increasing school system coverage may be extrapolated on the basis of recent experience. Similarly the rural impact of WOALP may be analyzed by the number who can be expected to pass standard literacy tests. In both cases the timing of the impact will differ and is one factor to include. In terms of the relevance of the assumptions of the analysis the feasibility of, say, a program aimed at six-week literacy training for inaccessible rural areas or differing language groups would have to be questioned. Simple extrapolation of current programs to areas with a number of widely differing characteristics may be extremely inaccurate.

As a consequence of assessing the effectiveness of the options the planner becomes aware of obstacles to implementation, consequences of different programs, and additional data needs. Hence step eight is to rate the alternatives according to legal and administrative obstacles and investigate the data needs required for final decisions.

Step nine is to analyze the possible results of the programs. This involves melding costs and effectiveness measures in an attempt to rank alternatives. At this point the three basic ingredients—objectives, costs, effectiveness—are combined as information to aid the decisionmaking process. This step involves not only assessment of the extent to which final objectives are attained and the costs of doing so, but also the added gains toward the measurable objectives that can be made at each level of program option and the added costs that would be incurred.

14

Outcomes And Evaluation

Many nations presently are engaged in improving their policymaking and program performance capabilities. Evaluation—or the appraisal of the effectiveness of programs—is a part of that endeavor. Distinguished from policy analysis in that it is after the fact, evaluation can provide important data to improve analysis of alternative policies. In drawing on existing experiences evaluation provides a guide to the "best" options.

Evaluation can range narrowly from project monitoring through various stages up to a broad concern for overall policy impact. Overall policy impact is concerned with the general system in which a policy is carried out. When evaluation is directed so broadly it is necessary to identify components of the general system and assess the contribution of each. Policy impact evaluation in education, for example, has identified and assessed the role of the family, peer groups, neighborhood, and formal, as well as less traditional, education. (The Coleman Study and its subsequent reviews are examples of broad policy impact assessment.) Another example of evaluation directed at an overall policy impact would be one addressed to the questions of whether educational expenditure is achieving the purposes for which vast amounts of money are being spent.

Policy impact evalution is heavily dependent upon a multidisciplinary body of knowledge. For instance, if the issue is learning, researchers would have to draw from many fields of study: learning theory, teaching processes, and relationships of school to family, to neighborhood, to play and to work. Such evaluation—as one might well expect—is difficult due to our poor understanding of the outcomes of behavioral relationships and interactions.

Differential effectiveness assessment represents another level of evaluation which addresses itself to such questions as: Are some strategies more effective than others in achieving the purposes sought? Or does a particular program work better or cost less for one group of persons than another?

A third level, project monitoring, examines specific activities rather than overall developmental programs. Much of the emphasis of project monitoring is on getting reliable data about the project both as to cost and activities.

However broadly or narrowly directed evaluation remains essentially a process for measuring progress and answering questions such as: Is the activity, program, or policy a success? Those qustions cannot be answered unless there are yardsticks available by which to assess progress. Without such "yardsticks of success" evaluation is impossible; with them, it is encouraged. Furthermore it is not a simple matter to determine what the outcomes of any activity are. Most programs have multiple purposes. What then to evaluate is a hard issue, and what is more not all identifiable outcomes can be evaluated, for some are more difficult to get data on, and in many cases evaluations are costly.

It is often difficult to choose among objectives and the measures of outcome appropriate to those objectives. Agreement on outcomes is likely to be difficult. A workable formulation of outcomes that specifies proxy measures can be achieved without reaching such a full accord. It is easier to reach this point when a number of indicators are used rather than one. For each the question to be addressed is: What difference did the program make?

Designing evaluation studies in terms of selected outcomes is set forth below as a series of requirements taken from a U.S. draft report (General Accounting Office, 1974):

- Program evaluation studies should isolate and specify the various elements and phases of the program studied and the several elements and phases of its physical, biological, general social, organizational, and economic environments that affect its operation and their results, and should state fully the causal and other relationships among them.
- Functional relationships of benefit, effectiveness or output variables should be hypothesized in terms of resources and conditions and constraints imposed; the evaluation should seek to accept or reject these hypotheses and quantify them.
- Program models should be utilized when they can generate more useful, comprehensive, or economical results to the fullest extent feasible; the evaluator is to choose or devise the most relevant model and use it with rigor. Such models will depict the flow of resource inputs into activities that cause results as conditioned by environmental and institutional constraints.
- Statistical estimation procedures should be designed to quantify the variables and model inputs from feasible sources of known validity; if validity is in doubt and either time or resources do not permit new data gathering, the risk of erroneous conclusions should be weighed before proceeding.
- Sampling procedures should conform to recognized published standards that can be documented and referenced or the departure from such standards should be explicitly and clearly explained and justified.

- Assumptions used at any stage of the inquiry and in all constituent analyses of the evaluation study should be realistic, clearly specified, and to the extent possible verified as to accuracy. Any recognized bias in the assumptions should be identified and clearly stated. The evaluator should continuously reassess assumptions with which the study started and change them and recycle the analysis when this appears needed.
- The assessment should be sufficiently broad in scope as to provide findings that can satisfy lines of inquiry required to understand the results of the educational services.

One more point should be added about the design of evaluation studies. Evaluations tend often to be static in their design. While structuring a system to study and define output measures for current programs, it is important to build into that structure the capability for change and evolution. As more becomes known new program goals and objectives will be defined and old ones abandoned. Changes may not come in large quantum jumps, but rather in small realignments, resulting from new knowledge, new perceptions, and a broader understanding of the problem and its potential solutions. Evaluation capabilities must improve and be redeveloped over time as new and different education programs enter the picture. We must view evaluation as a kinetic, dynamic function rather than as the application of old, static, or perhaps obsolete criteria to assess programs that are constantly developing and changing.

Once the outcomes to be evaluated are selected and once the evaluation program is designed one must begin collecting data. This reservoir of data should be as complete as necessary to support a full analysis. Existing records (maintained so as to be suitable); surveys of target groups, administrators, or others interested; periodic reports (referenced to records or other sources); and other research and program evaluation study reports should all be utilized. Ideally the data should reach the varied parameters of the program, including side-effects, by-products, and secondary results. But realistically, this is often impossible, for much of the data collected is likely to be proxy, that is, partial or symptomatic information standing in for a general category of results more difficult to quantify completely. For example in assessment of earnings in the computation of investment returns a proxy—"average earnings"—often stands in for marginal wages; and wages at the margin may be for all classes of workers rather than workers with the specific characteristics of those under study.

In any event whether "proxy" or "full-parameter" information is used, only those statistical procedures that are well known and readily interpreted should be utilized for collecting and analyzing any quantified information. Any uncommon statistical techniques employed should be clearly referenced to authorities who developed or tested them.

Another important part of evaluation is project monitoring. Project monitoring is intended to make sure that records of program inputs, processes, and outputs are available. As indicated earlier it is mainly an administrative function, but interproject comparisons can yield experimental findings. The extent to which such findings can be drawn from the variation among projects depends upon objective observation and the capacity to find true controls. That is to say when we seek an answer to the question, what difference does the project make, we have to ask: Differences compared to what? And the basis for comparison must be some project that provides the function of an experimental type of control.

Indeed it has been advocated that planned variations be built into projects to serve as control groups for the purposes of experimentation. In any event the normal process of attempting to judge "success" generates experimentation, for in the process of so doing it becomes clear that much necessary knowledge is lacking about program development and the way in which outcomes can be obtained. While experimentation in social programs is difficult, evaluation moves in that direction because of difficulties of telling whether or not programs have had the intended effects.

There are, then, minimal requirements of a comprehensive project evaluation if it is to provide useful policy guidance over the program: a) the evaluation system must provide for autonomy of evaluation efforts; and b) evaluation systems must provide evaluators with compliance control.

The vested interests—biases, prejudices, short-sightedness and honest convictions of those persons closely associated with programs—may hamper objective evaluation if they are involved in the evaluation process. Administrators and staff establish biases that severely limit their ability as a group to contribute to evaluation efforts. For this reason there is good cause to be somewhat skeptical of data or in-house evaluations provided by persons working within specific programs.

Evaluation—like all other human activities—is constrained by limited finances. Independent investigations—though ideal—cost more in terms of time and personnel. Evaluation must, therefore, depend partly—often largely—on those persons working within the project to provide essential information and perspective. Given the incentives noted above for persons working in particular programs or projects to bias or block evaluation efforts, the agency charged with evaluation must have the authority to obtain information. But compulsory compliance forced upon project administrators and staff will probably not yield useful results in terms of subsequent action. Half-hearted or counterproductive efforts forced out of program or project staff is not the goal of compliance control. The knowledge that deliberate lack of a cooperative effort may result in an evaluation report biased against the program should be incentive for program personnel to provide necessary data. The goal of the evaluating agency should be a cooperative effort with a substantial and fair hearing of the ideas and opinions of those within the program. All this suggests that the structure of the evaluation function must be hierarchical, involving persons from the project level on up with adequate shares of program data accessibility and evaluation objectivity.

THE USES OF EVALUATION AND OUTCOME CRITERIA

Evaluation and outcome data have a number of uses:

1. *Defining program potential*—evaluation of past performances and goals often deflates unwarranted optimism and gives planners a better idea of what is feasible within certain resource constraints.

When this consequence of outcome measurements is fully reflected in policy decisions it will be clear that the generalities of earlier days have been abandoned and in their place there is far greater reality and specificity borne of the greater knowledge acquired of program operations in terms of measures of "success" and "results." For example, through evaluation of education we have defined better the role of the family, neighborhood, and peer groups in learning, and consequently have obtained more realistic views about what can be accomplished during the limited time spent in school.

2. *Targeting resources*—evaluation facilitates a focus of policy on critical needs. When outcome data question disparities in allocation of resources new emphasis is given to marshaling those resources where they can meet the most pressing need more effectively.

3. *Encouraging program change*—outcome data based on established criteria, moreover, make it possible to deal with activities and programs in a function and changing system. The assessment of results encourages monitoring of projects to determine whether they are on target; it also paves the way for subsequent changes in direction by revealing weaknesses and barriers.

4. *Facilitating cost comparisons*—the criteria (once available in measurable form) permit analysis of costs per unit for a defined result. Unless program output can be equated from one activity to another—or one program to another—cost comparisons are not very meaningful. Standardization of "units" is essential to such cost comparisons.

5. *Better understanding of program interaction*—derived from outcome assessment too is a better awareness of the relationships of programs and activities to human behavior. Evaluation also brings a better understanding of the problem of duration and timing in public programs and activities. For instance we have begun to give more emphasis to the length of time required to produce a given output with the resource available. And for learning, the question of optimum timing is becoming more vital.

6. *Encouraging a "numbers game"*—measurements that are incorporated in resource and program decision lend themselves to abuse. The more important the numbers become, the greater the incentive for abuse. Aside from inaccurate reporting—which may be of less importance—one

7

of the aspects of such abuse of the statistics from evaluation of program policies is that results are focused to yield the most favorable number result. If grade-repeaters in school are an index of a problem in teaching and the fewer repeaters the better, automatic promotion of even poor students becomes a policy. Or if the number of graduates is deemed the measure of success, graduation of nearly all is the response. The changes in numbers alter the criteria count but not the knowledge or skill of young persons in the nation.

Measurements of the objectives of any program tend to take on a dynamics of their own to which policy officials respond, so it is easy to understand how the abuse of statistics and evaluation tends to alter the very objectives themselves. Pressures to succeed, by whatever criteria, should not be discounted and certainly not ignored.

7. *Evaluation as an educational tool*—we earlier reviewed briefly outcome measures as educational devices. Evaluation using outcome criteria is useful as a guide to teaching, curriculum formulation, and to administrators who carry responsibility for assuring that the criteria used are understood and reflective of teaching needs. In the case of programmed instruction detailed outcome objectives are fundamental to the design of the programs. In some instances individual student learning on a tutorial basis is formulated in terms of essentially a contract between student and teacher. Rating of progress toward specific goals is an integral part of this educational method.

Before concluding, it is important to point out at least three warnings concerning the use of evaluation studies.

First, while evaluation with its requirements for specificity of criteria of assessment offers great promise, it lends itself to premature political conclusions and manipulations. Politicians may draw too hastily upon evaluation findings.

Second, beginning efforts at evaluation have to be viewed as beginnings. A single evaluation study, no matter how well carried out, is not a sufficient guide for a *major* policy decision.

Third, learn when not to use evaluation's findings. A dean of the evaluation profession recently stated in discussion of the negative findings of the Westinghouse-Ohio University evaluation of Head Start, "In this case I think that the testimony of the parents who had children in school earlier and then had children in Head Start programs—their warm enthusiasms—were more valid indicators of the impact of Head Start than the stacks of computer output. I am glad that the political system disregarded a very impressive statistical analysis."

For those educational systems in which participation of groups and families is built into decisionmaking about programs and resource allocation, output measurements serve an added purpose: namely, to give meaning to such participation. A system of measuring educational outcomes also improves consumer awareness about education. Information available on the various outcomes of various types of educational programs and schools permits a more informed choice.

Output measurement thus permits a political arithmetic on educational systems. It is an arithmetic of human resource development for growth, for improved quality of living and for enhanced individual enjoyment on the one hand, and preservation of scarce resources on the other.

Once established as a routine of educational and public policy, outcome measurements will have important feedbacks. It is imperative that those feedback results be of a positive nature and that they help the responsible officials to build sounder policies and projects that will improve the efficiency and effectiveness of institutional arrangements for instilling knowledge, skills and values.

Experience in the United States and Ethiopia points to both the large potential in application of outcome measurements and the need for caution to ensure that new outcome indicators measure education's development potential more fully. In conclusion, what is needed is a better understanding of what education can do, what is within its capacity, and what may best be done elsewhere.

Bibliography

1. Adelman, I. and Morris, C.T. *Society, Politics and Development*. Baltimore: Johns Hopkins Press, 1967.
2. Anderson, C.A. "The Social Context of Educational Planning." UNESCO: IIEP, *Fundamentals of Educational Planning*. Paris, France: UNESCO, 1967.
3. Averch, H., et al. "How Effective is Schooling?" Rand Corporation, Santa Monica, California. March 1972, pp. 1-222.
4. Bachman, G. *Youth in Transition*. Vol. II: *The Impact of Family Background and Intelligence of Tenth Grade Boys*. Ann Arbor, Michigan: Institute for Social Research, 1970.
5. Bloom, B.S., et al, *Taxonomy of Educational Objectives, Handbook I: Cognitive Domain*. New York: David McKay Co., Inc., 1954 (copyrighted 1956).
6. Brembeck, Cole S. "Education and Social Change." *Education and the Development of Nations*. Edited by Hanson, John W. and Brembeck, Cole S. New York: Holt, Rinehart and Winston, 1966.
7. Bryant, E.C., Glazer, Ezra, and Hansen, Morris H. "Association Between Educational Outcome and Background Variables: A Review of the Literature." Prepared for the National Assessment of Educational Progress under the support of the U.S. Office of Education, November 1973.
8. Buros, Oscar K. (ed.). *The Mental Measurement Yearbook*, 7th Ed., Highland Park, New Jersey: Gryphon Press, 1972.
9. The Center for the Study of Evaluation. *Elementary School Test Evaluations*. Los Angeles, California: The Center for the Study of Evaluation of the University of California, 1970.
10. ________. *Preschool Kindergarten Test Evaluation*. Los Angeles, California: The Center for the Study of Evaluation of the University of California, 1970.
11. Coleman, James. *Education and Political Development*. Princeton, New Jersey: Princeton University Press, 1967.
12. Comber, L.C. and Keeves, J.P. *Science Education in Nineteen Countries*. New York: Wiley-Halsted Press, 1973.
13. "Commission on Primary Education for the Conference in Education and Scientific and Technical Training." Nairobi, Kenya.
14. Coombs, Philip H. and Hallak, Jacques. *Managing Educational Costs*. London: Oxford University Press, 1972.
15. Coopersmith, Stanley. *Self Esteem as a Determinant of Selective Recall and Repetition*. (Doctoral thesis, Cornell University, 1958.)
16. Crowne, Douglas P. and Marlowe, David. *The Approach Motive: Studies in Evaluation Dependence*. New York. John Wiley & Sons, 1974.

17. Dawson, J.L.M. "Traditional Values and Work Efficiency in a West African Mine Labour Force." *Occupational Psychology*, 39:209-218, 1963.

18. Denison, Edward. "Measuring the Contribution of Education to Economic Growth," *The Residual Factor and Economic Growth*. Paris: OECD, 1964.

19. Diez-Hochleitner, R. "Policies and Practices in Educational Planning," Lecture-Discussion Series No. 10. Paris: UNESCO, IIEP, 1967.

20. El Koussey, Abdel Aziz Hamid. "For a Self Criticism of Education in the Arab Countries." *Prospects: Quarterly Review of Education*, UNESCO, 3 (1), Spring 1973.

21. Encyclopedia Britannica, 1974. Chicago, Toronto, London, Geneva, Sydney, Manila, Seoul, Johannesburg, Tokyo: William Benton Publishers.

22. Ethiopia, Department of Labor, Ministry of National Community Development and Social Affairs. *An Assessment of Ethiopia's Manpower Requirements and Resources for Economic Development 1961-70 E.C.* Directory of National Community Development and Social Affairs. Addis Ababa: 1970.

23. ________. Report of the Education Sector Review. *Education: Challenge to the Nation*. Addis Ababa, Ethiopia: August 1972.

24. ________. Institute of Development Research. "The Ghimbi Rural Survey." Addis Ababa, Ethiopia: 1973.

25. ________. Ministry of Education. *Reports*. Addis Ababa, Ethiopia: 1970.

26. ________. Ministry of Education and Fine Arts/UNOP/UNESCO Work Oriented Adult Literacy Project. *Functional Literacy in Ethiopia—A Pilot Experiment*. Addis Ababa, Ethiopia: 1972.

27. ________. Planning and Programming Office, Ministry of Education and Fine Arts, *Program Budgeting. A Suggestion for Improving the Ministry's Budgeting Procedure*. Addis Ababa, Ethiopia: 1972.

28. ________. Planning Commission Office. "Economic Return to Investment in Ethiopian Education." (mimeo.) Addis Ababa, Ethiopia: 1973.

29. Fassil, G. Kiros, et al. "Financing of Education." *ESRC/5, Report of Education Sector Review*. Addia Ababa, Ethiopia.

30. Faure, Edgar, et al. *Learning to Be*. Paris, France: Harrap, UNESCO, 1972.

31. Fischer, John. "Indonesia." *Education and the Development of Nations*. Edited by Hanson, John W. and Brembeck, Cole S. New York: Holt, Rinehart and Winston, 1966.

32. Gage, N.L. (ed.). *Handbook of Research on Teaching*. Chicago: Rand McNally and Co., 1963.

33. Hong, Sung Chick. *The Intellectual and Modernization*. Seoul, Korea: Research Institute, Korea University, 1967.

34. Horton, Robin. "African Traditional Thought and Western Science I." *Africa*, 37(1):50-71.

35. ________. "African Traditional Thought and Western Science II." *Africa*, 37(2):155-187.

36. Inkeles, Alex. "Making Men Modern: or the Causes and Consequences of Individual Change in Six Developing Countries." *American Journal of Sociology*, 72(2):208-225, 1969.

37. International Council on Educational Development. *Attacking Rural Poverty*, 1973.

38. Jallade, Jean-Pierre. *Public Expenditures on Educational and Income Distribution in Colombia.* World Bank Staff Occasional Papers, No. 18, Washington: International Bank for Reconstruction and Development, 1974.

39. Jencks, Christopher. *Inequality: A Reassessment of the Effect of Family and Schooling in America.* New York: Basic Books, 1972.

40. Kajubi, W. Senteza. "New Directions in Teacher Education in East Africa." *International Review of Education,* 17(3):197-210, 1971.

41. Krathwohl, D.R. *Taxonomy of Educational Objectives Handbook II: The Affective Domain.* New York: David McKay Co., Inc., 1964.

42. Krug, Robert E and Schwarz, Paul A. *Ability Testing in Developing Countries: A Handbook of Principles and Techniques.* New York: Praeger Publishers, 1972.

43. Levenson, H. "Distinctions within the Concept of Internal External Control: The Development of a New Scale." Paper presented at the American Psychological Associaton, Hawaii, 1972.

44. Maclure, J. Stuart. "Curriculum Innovation in Practice: Canada, England, Wales, U.S." Third International Curriculum Conference, Oxford. Her Majesty's Stationery Office, 1968.

45. McClelland, David. *The Achieving Society.* Princeton, New Jersey: D. Van Nostrand Company, Inc., 1961.

46. ________, and Winter, David G. *Motivating Economic Achievement.* New York: The Free Press, 1969.

47. Manaster, Guy J. and Havighurst, Robert J. *Cross-National Research, Social-Psychological Methods and Problems.* Boston: Houghton-Mifflin and Co., 1972.

48. Martin, F. *The Creative Organization of Positive Experiences.* Lawrenceville, Georgia: Gwinant County Schools, 1972.

49. Mayjeske, George, et al. *A Study of Our Nation's Schools.* Washington, D.C.: Office of Education, Department of Health, Education, and Welfare.

50. Mosteller, Frederick and Moynihan, David P. (eds.) *On Equality of Educational Opportunity.* New York: Vintage Publishers, 1972.

51. Mushkin, Selma J. and Pollack, William. "Analysis in a PPB Setting." *Economic Factors Affecting the Financing of Education.* Ed. by Johns, R.L., Goffman, I.J., Alexander, K. and Stoller, D.C. Washington, D.C.: National Educational Finance Project, 1970.

52. Nowicki, S. and Walker, C. "Achievement in Relation to Locus of Control: Identification of a New Source of Variance," *Journal of Genetic Psychology* (in press).

53. Ord, I.G. "Testing for Educational and Occupational Selection in Developing Countries—A Review." *Occupational Psychology,* 46(3):125-182, 1972.

54. Platt, William J. "Policy-Making and International Studies." Paper presented at Conference on Educational Achievement, Harvard University, November 1973.

55. Purves, Alan C. *Literature Education in Ten Countries.* New York: Wiley-Halsted Press, 1973.

56. Razik, T.A. *Systems Approach to Teacher Training and Curriculum Development: The Case of Developing Countries.* Paris: UNESCO, 1972.

57. Robinson, John P. *Measures of Political Attitudes*. Ann Arbor, Michigan: Survey Research Center, Institute for Social Research, 1968.

58. ________. *Measures of Occupational Attitudes and Occupational Characteristics*. Ann Arbor, Michigan: Survey Research Center, Institute for Social Research, 1968.

59. ________. *Measures of Social Psychological Attitudes*. Ann Arbor, Michigan: Survey Research Center, Institute for Special Research, 1973.

60. Schiefelbein, Ernesto. "Constraints to Change in Traditional Educational Systems: Lessons from Chile." *Interchange*, 2(4), 1971.

61. Singh, B.W., Teklehaiman, Sestor and Abbaye, Laeke. *Relationship Between Functional Literacy and Economic Development in Chilalo Awraja (Armin Province)*. Addis Ababa, Ethiopia: UNESCO Work Oriented Adult Literacy Project, 1973.

62. Soares, L.M. and Soares, A.T. "Interpersonal Perception of Disadvantaged Children." *Proceedings*, 79th Annual Convention of the American Psychological Association, 1971.

63. Swart, M.S. and Swart, R.C. "Self Esteem and Social Personal Orientation of Indian 12 and 18 Year Olds." *Psychological Reports*, 1970, 107-115.

64. Thias, Hans Heinrich and Carnoy, Martin, *Cost Benefit Analysis of Education: A Case Study of Kenya*. Washington, D.C.: International Bank for Reconstruction and Development, 1972.

65. Thorndike, Robert L. *Reading Comprehension Education in Fifteen Countries*. New York: Wiley-Halsted Press, 1973.

66. Tickton, Sidney G. "Instructional Technology in the Developing World." *Educational Broadcasting Review*, 6(2), April 1972.

67. UNESCO. *Book Development in Africa: Problems and Perspectives*. Reports and Papers on Mass Communications #56. Paris: UNESCO, 1968.

68. ________. *Educational Planning: A World Survey of Problems and Prospects*. Paris: UNESCO, 1970.

69. ________. UNESCO Yearbook, 1970, Paris: UNESCO, 1970, p. 31.

70. Zack, A.M. and Mohammed, Duri. *Ethiopia's High-Level Manpower Analysis and Projection*. Addis Ababa, Ethiopia: Mimeo., 1964.

71. Ziller, R.C. and Ziller, L.H. "Political Personality." *Proceedings*, 79th Annual Convention of the American Psychological Association, 1969.

Index

Abidjian Conference (1961), 140
Ability, innate, 19, 53, 165
Accion Cultural Popular (ACPO), 96
Achievement motivation, 87 88, 89
Addis Ababa Conference (1961), 140, 147
Administrators, 159-160
Affective variables, 5, 10, 26-28
African countries educational goals, 26, 79
Age-income profiles, 125-128
Agriculture (see rural)
Amharic proficiency test, 24, 25
Analytic techniques in developing countries, 152-153
Anomie, 33, 34
Aptitude and achievement testing, 19-21

Behavioral outcome measures, 91, 92
Bloom, Benjamin, 22, 29
Buros' Mental Measurement Yearbook, 33

Cambridge Overseas Examination Syndicate, 13
Center for the Study of Evaluation, University of California, 33
Ceylon, 152
Chilalo Agricultural Development Unit (CADU), 96, 100, 102, 103, 121, 122
Chile, 153
China, 56, 57
Coff-Douglas production function, 118, 119
Cognitive factors, 5, 6, 10, 22, 30, 31
Coleman study, 54, 76
Colombia, 96, 137, 138
Cost benefit analysis 121-137, 149, 150
Criterion reference test, 21, 22
Cross-cultural studies, 36
Crown-Marlowe Scale, 34
Curriculum and outcome measures, 160-166; "hidden," 163-164

Department of Defense (U.S.) job tests, 110
Denison, Edward, 118-119
Diey-Hochleitner, 143

Education and economic growth, 2-3, 117, 118, 119; and personal development, 4-6; and productivity, 117, 128; and societal development, 8
Education Sector Review Commission, 79, 145-148
Economic Commission for Africa, 140
Educational Testing Service, Princeton, 33
Entrepreneurial activity measures, 87, 88
Ethiopia age-income variables, 125-127; Chilalo Agricultural Development Unit, 96, 100-104, 120, 121; Education Sector Review Commission, 79, 145-148; Haile Sellaisie I University, 16, 24, 26, 33, 40-44; Rural Survey, 91, 92; School Leaving Certificate Examinations, 12
Evaluation and outcome criteria uses, 182; design, 179-182; of educational outcomes, 153-156
Examinations (see tests)
External tests, 11

Fatalism, 86, 87

Gross National Product, 118, 119

Haile Sellaisie I University Testing Center, 16, 24, 26, 33, 40-44
Head Start, 183
Husen, Torsten, 63, 72, 73

Illiteracy, international comparisons, 59-62
Income as outcome measure, 125-128; redistribution, 137, 138
India, 46, 141-142
Indigenous measures, 76
Indonesia, 13-14
Institute for Social Research, University of Michigan, 33
Institute of Development Research, Haile Sellaisie I University, 91
Intergenerational impact of education, 136, 137
Internal rate of return (IROR), 128-133
International Association for the Evaluation of Education (IEA), 14, 15, 62-78
International Council on Educational Development, 95
International Labor Organization, 105
Investment outcomes, 128-137
Ivory Coast, 153

Jallade, Jean-Pierre, 137, 138
Job analysis, 106

Karachi Conference Plan, 141
Kenya, 95, 125, 127
Korea, 95

Language impact (in test design), 42-44
Literacy, 56, 57, 59-62
Literature testing (IEA), 66, 67
Locus of control, 26, 35-40

Macroeconomic measures of educational outcomes, 117
Manpower planning, 150, 151, 153, 167, 168
Mayeske, George, 54-55
Ministry of National Community Development and Social Affairs, Ethiopia, 105, 115
Minority group children (U.S.), 27
Mobile Trade Training Schools (Thailand), 95
Modernity indicators, 84, 85

Nairobi Conference (1968), 26, 79
National Case Study Questionnaire (IEA)
National Occupational Competency Testing Institute, 113-114
Need for achievement (nAch), 87, 88, 89
Net present discount value (NPDV), 128-132
Nigeria, 96
Non-academic skills, 26, 27
Non-cognitive outcomes, 164-165
Nonformal education, 95-103

Occupational testing, 105
Organization of American States (OAS), 142
Outcome measures as tool in development, 2-5; teacher training in, 157-159; use in curriculum, 160-166

Paris Conference (1962), 140
Philippines, 13
Planning process, schematic presentation, 154
Planning - programming - budgeting (PPB) system, 167-177; components, tools and processes of, 173
Political outcomes, 79-94
Productivity effect of education, 119, 120, 176
Project Talent, 54
Psycho-motor learning, 10

Quality of life, 6, 7, 8

Rate of Return Analysis, 149, 150, 151
Reading Comprehension Testing (IEA), 67-68
Recurrent education, 163
"Residual" approach, 118, 119
Resource allocation, 1, 9
Rural programs, 91, 95, 96, 176

Self-esteem, 26, 35-40
Senegal, 96
School Leaving Certificate Examination (ESLCE), 12
Science testing (IEA), 65, 66
Singapore Ministry of Education, 13
Social outcomes, 79-94
Socioeconomic Status (SES), 52-61
Standardized testing, 13-25, 48-49, 161, 162
Status Concern, 34
System Analysis, 167, 168, 169

Teacher tests, 45-51; training in outcome measures, 157-159
Tests, adaptation of, 15, 16, 85; anxiety factors in, 34, 86, 87; construction of, 14-17, 23; criterion referenced, 21, 22; external, 11; language constraints, 42-44; literature, 66, 67; reading, 67, 68; science, 65, 66; standardized, 13-25, 48-49; teacher, 45-51; trends in developing countries, 45-46; validity of, 23, 24; work-skill, 105-115
Thailand, 95, 152
Thematic Apperception Test (TAT), 86
Tokyo Conference (1962), 141

Uganda, 152
UNESCO, 46, 59-62, 140

Validity (for tests), 23, 24
Vietnam, 13
Village Polytechnics (Kenya), 95
Vocational training, 105

Work Oriented Adult Literacy Project, 96, 100-104, 120, 121
Work skill tests, 105-115